

# ESTADÍSTICA PARA QUÍMICA ANALÍTICA

Segunda edición

J.C. Miller • J.N. Miller



ADDISON-WESLEY IBEROAMERICANA

# ESTADÍSTICA PARA QUÍMICA ANALÍTICA

**Segunda Edición**

**J. C. MILLER**

*Wyggeston and Queen Elizabeth I College, Leicester*

y

**J. N. MILLER**

*Decano de ciencias puras y aplicadas  
Loughborough University of Technology*

Versión en español de

**Roberto Izquierdo Hornillos**

*Profesor de Química Analítica  
Universidad Complutense de Madrid*

y

**Carlos Maté Jiménez**

*Profesor de Estadística  
ICAI, Universidad Pontificia Comillas*

REVISIÓN DEL CAPÍTULO  
12

• CAPÍTULO 12. ANÁLISIS DE  
REGRESIÓN

Con la colaboración de

**Santiago Capella Vizcaíno**

*Universidad Nacional Autónoma de México*



**ADDISON-WESLEY IBEROAMERICANA**

Argentina • Brasil • Chile • Colombia • Ecuador • España  
Estados Unidos • México • Perú • Puerto Rico • Venezuela

Esta versión en español es la única autorizada.

**Portada:** Ricardo Arriaga

© 1993 por Addison-Wesley Iberoamerican, S.A.  
Wilmington, Delaware, E.U.A.

Impreso en Estados Unidos. *Printed in U.S.A.*

ISBN 0-201-60140-0

1 2 3 4 5 6 7 8 9 10-AL-97 96 95 94 93

# Índice general

<b>Prefacio</b> . . . . .	ix
<b>Prefacio a la segunda edición</b> . . . . .	xi
<b>Glosario de símbolos</b> . . . . .	.xiii
<b>1 Introducción</b>	
1.1 Problemas analíticos . . . . .	1
1.2 Errores en el análisis cuantitativo . . . . .	2
1.3 Tipos de errores . . . . .	3
1.4 Errores sistemáticos y aleatorios en el análisis volumétrico . . . . .	7
1.5 El manejo de errores sistemáticos . . . . .	11
1.6 Planificación y diseño de experimentos . . . . .	14
1.7 Calculadoras y computadores en los cálculos estadísticos . . . . .	15
Bibliografía . . . . .	17
Ejercicios . . . . .	18
<b>2 Errores en el análisis clásico. Estadística de medidas repetidas</b>	
2.1 Media y desviación estándar . . . . .	20
2.2 Distribución de errores . . . . .	22
2.3 La distribución muestral de la media . . . . .	25
2.4 Límites de confianza de la media . . . . .	28
2.5 Presentación de resultados . . . . .	31
2.6 Otros usos de los límites de confianza . . . . .	32
2.7 Propagación de errores aleatorios . . . . .	32
2.8 Propagación de errores sistemáticos . . . . .	36
Bibliografía . . . . .	37
Ejercicios . . . . .	37

<b>3 Pruebas de significación</b>	
3.1 Introducción	40
3.2 Comparación de una media experimental con un valor conocido	40
3.3 Comparación de las medias de dos muestras	42
3.4 La prueba $t$ por parejas	44
3.5 Las pruebas de una y dos colas	46
3.6 La prueba $F$ para la comparación de desviaciones estándar	47
3.7 Valores anómalos	49
3.8 Análisis de la varianza	51
3.9 Comparación de varias medias	52
3.10 La aritmética de los cálculos ANOVA	55
3.11 La prueba chi cuadrada	57
3.12 Prueba de la normalidad de una distribución	58
3.13 Conclusiones a partir de las pruebas de significación	61
Bibliografía	63
Ejercicios	63
<b>4 Control de calidad y muestreo</b>	
4.1 Introducción	67
4.2 Muestreo	67
4.3 Estimación y separación de varianzas utilizando ANOVA	69
4.4 Estrategia de muestreo	70
4.5 Ensayos de colaboración: introducción	71
4.6 Representaciones gráficas de dos muestras	72
4.7 La preparación de un ensayo de colaboración	73
4.8 Cálculos en ensayos de colaboración	75
4.9 Diagramas de control	78
Bibliografía	84
Ejercicios	84
<b>5 Errores en análisis instrumental; regresión y correlación</b>	
5.1 Análisis instrumental	87
5.2 Gráficas de calibración en análisis instrumental	88
5.3 El coeficiente de correlación momento-producto	90
5.4 La recta de regresión de $y$ sobre $x$	95
5.5 Errores en la pendiente y ordenada en el origen de la recta de regresión	96
5.6 Cálculo de una concentración	98
5.7 Límites de detección	100
5.8 El método de las adiciones estándar	103
5.9 El uso de rectas de regresión para comparar métodos analíticos	106
5.10 Rectas de regresión ponderadas	110
5.11 Regresión curvilínea	114
Bibliografía	119
Ejercicios	119
<b>6 Métodos no paramétricos y de ejecución rápida</b>	
6.1 Introducción	122

6.2 La mediana	123
6.3 La prueba de signos	124
6.4 La prueba de rachas de Wald-Wolfowitz	126
6.5 Pruebas basadas en el recorrido de los resultados	127
6.6 La prueba de rangos y signos de Wilcoxon	129
6.7 La prueba de suma de rangos de Wilcoxon y métodos relacionados	131
6.8 Pruebas no paramétricas sobre más de dos muestras	135
6.9 Correlación ordinal	137
6.10 Métodos de regresión no paramétricos	139
6.11 La prueba de bondad del ajuste de Kolmogorov	141
6.12 Conclusiones	145
Bibliografía	145
Ejercicios	145
<b>7 Diseño experimental, optimización y reconocimiento de pautas</b>	
7.1 Introducción	147
7.2 Aleatorización	148
7.3 Experimentos con bloques	148
7.4 ANOVA de dos factores	149
7.5 Cuadrados latinos	152
7.6 Diseños anidados y de clasificación cruzada	153
7.7 Interacción	153
7.8 El diseño factorial frente al de un factor cada vez	159
7.9 Diseño factorial y optimización	159
7.10 Incapacidad del método de un factor cada vez para lograr el óptimo	163
7.11 Método de la máxima pendiente o del gradiente	164
7.12 Optimización por el método simplex	168
7.13 Reconocimiento de pautas	171
7.14 Métodos de aprendizaje supervisados	173
7.15 Métodos de aprendizaje no supervisados	176
Bibliografía	178
Ejercicios	179

Soluciones a los ejercicios	181
-----------------------------	-----

Apéndice 1 Resumen de pruebas estadísticas	196
--	-----

Apéndice 2 Tablas estadísticas	199
--------------------------------	-----

Índice de materias	210
--------------------	-----



## Prefacio

Añadir otro volumen a los numerosos textos de estadística pudiera parecer una tarea sin excesivas garantías de éxito; no obstante, la realidad es que por desgracia muchos científicos altamente calificados ignoran hasta los métodos estadísticos más elementales. Es más asombroso aún que los químicos analíticos, quienes practican una de las ciencias más cuantitativas de todas las existentes, no estén más prevenidos que otros contra este mal peligroso, pero sin lugar a dudas curable. Por lo tanto, es de esperar que este libro pueda beneficiar a los científicos analíticos que deseen diseñar y realizar sus experimentos en forma correcta, y extraer tanta información de los resultados como puedan. Así mismo, pretende atraer al creciente número de estudiantes que se especializan en química analítica, y a los que utilizan métodos analíticos en forma rutinaria en el trabajo de laboratorio.

Existen otras dos razones que nos han alentado a escribir este libro. Una es el enorme impacto de la microelectrónica, en forma de microcomputadores y calculadoras de bolsillo, sobre la estadística: mediante estos aparatos, los científicos activos han resuelto los problemas que planteaban los procedimientos analíticos difíciles. La segunda es el rápido desarrollo de los nuevos procedimientos “quimiométricos”, incluyendo el reconocimiento de pautas, optimización, técnicas de filtrado numérico, simulaciones, etc., todos ellos factibles gracias a las facilidades que aportan los computadores. En el último capítulo de este libro se intenta proporcionar al lector cuando menos una introducción del potencial de algunos de estos métodos estadísticos más novedosos. Sin embargo, no se ha incluido ningún programa de computador en el libro; esto se debe, en parte, a las dificultades de presentar programas que puedan ejecutarse en todos los tipos conocidos de microcomputadores, y en parte a la existencia de una cantidad considerable de libros y programas de computador adecuados y accesibles al público.

La disponibilidad de este enorme potencial de los computadores hace que el científico, naturalmente, aplique métodos estadísticos de manera racional y correcta. Para limitar la extensión de este libro, y recalcar sus aspectos prácticos, no hemos intentado mostrar con detalle las bases teóricas de las pruebas estadísticas descritas. No obstante, hemos intentado aclarar al analista cuáles son las pruebas apropiadas para los tipos de problemas que probablemente encontrará en el laboratorio. En el texto se incluyen ejemplos resueltos y,

al final de cada capítulo, ejercicios para el lector. Muchos de ellos se basan en los datos proporcionados por trabajos de investigación publicados en *The Analyst*. Nuestro profundo agradecimiento al editor Phil Weston, por permitirnos hacer uso de su distinguida revista. También agradecemos a nuestros colegas, amigos y familia su paciencia durante la preparación del libro; por otro lado en los apéndices se citan individualmente las fuentes de las tablas estadísticas; al editor de la serie, doctor Bob Chalmers; y a nuestros editores por su eficiente cooperación y consejos.

J. C. Miller  
J. N. Miller

## Prefacio a la segunda edición

La gran aceptación que ha tenido la primera edición de este libro nos ha animado a preparar una segunda edición revisada y ligeramente ampliada. Agradecemos en especial a quienes han indicado errores de la primera edición y nos sugirieron posibles mejoras. El profesor D. Thorburn Burns nos ha ayudado mucho al señalarnos los errores; también agradecemos al profesor J. E. Colbeck por sus comentarios constructivos acerca de la nomenclatura.

Los principales cambios introducidos en esta edición forman un nuevo capítulo (4), en el cual se trata por primera vez los ensayos de colaboración y se cubre en gran medida el muestreo y los diagramas de control. Se añadió material sobre errores de tipo 1 y tipo 2; el uso secuencial de las pruebas de significación; los métodos no paramétricos, y las gráficas de calibración no lineales. Por último, se amplió el estudio del reconocimiento de pautas y del diseño factorial, y se incorporaron soluciones más detalladas a algunos de los problemas.

De nuevo, agradecemos a nuestros editores; los editores de la serie, los doctores Bob Chalmers y Mary Masson; y Phil Weston, editor de *The Analyst*, su ayuda y cooperación adicionales para la producción de esta edición.

J. C. Miller  
J. N. Miller

## Glosario de símbolos

$a$	— ordenada en el origen de la recta de regresión
$b$	— pendiente de la recta de regresión
$c$	— número de columnas en el ANOVA de dos factores
$C$	— término de corrección en ANOVA de dos factores
$F$	— razón de dos varianzas
$F_R$	— razón de intervalos
$h$	— número de muestras en el ANOVA de un factor
$\mu$	— media aritmética de una población
$M$	— número de signos menos en la prueba de rachas de Wald-Wolfowitz
$n$	— tamaño muestral
$N$	— número de signos más en la prueba de rachas de Wald-Wolfowitz
$N$	— número total de mediciones en ANOVA de dos factores
	— número de grados de libertad
$P(r)$	— probabilidad de $r$
$Q$	— $Q$ de Dixon, utilizada para pruebas de valores anómalos
$r$	— coeficiente de correlación momento-producto
$r$	— número de filas en ANOVA de dos factores
$\rho$	— coeficiente de correlación de intervalos de Spearman
$s$	— desviación estándar de una muestra
$S_{y \cdot x}$	— desviación estándar de los residuos de $y$
$S_b$	— desviación estándar de la pendiente de la recta de regresión
$S_a$	— desviación estándar de la ordenada en el origen de la recta de regresión
$S_{(y/x)w}$	— desviación estándar de los residuos de $y$ de rectas de regresión ponderadas
$S_{\hat{x}(y)}$	— desviación estándar de los valores de $x$ estimados utilizando rectas de regresión
$S_B$	— desviación estándar de un blanco
$S_{\hat{x}(E)}$	— desviación estándar de un valor de $x$ extrapolado
$S_{\hat{x}(0w)}$	— desviación estándar de un valor de $x$ estimado utilizando rectas de regresión ponderadas
$\sigma$	— desviación estándar de una población

$\sigma_0^2$	—	varianza de una medición
$\sigma_1^2$	—	varianza muestral
$t$	—	cantidad utilizada en los cálculos de los límites de confianza y en pruebas de significación de la media (véase la Sec. 2.4)
$T$	—	gran total en el ANOVA
$T_d$	—	cantidad utilizada en pruebas de significación de la media (véase la Sec. 6.5)
$T_1$ y $T_2$	—	estadísticos utilizados en la prueba de la suma de rangos de Wilcoxon
$w$	—	intervalo
$w_i$	—	peso dado a un punto en la recta de regresión
$\bar{x}$	—	media aritmética de una muestra
$\bar{x}_0$	—	valor de $x$ estimado utilizando rectas de regresión
$x_E$	—	valor de $x$ extrapolado
$\bar{x}_w$	—	media aritmética de valores de $x$ ponderados
$\chi^2$	—	cantidad utilizada en contrastes de bondad del ajuste
$y$	—	predicción del valor de $y$ por la recta de regresión
$\bar{y}_w$	—	media aritmética de los valores de $y$ ponderados
$y_B$	—	señal del blanco
$z$	—	variable normal estándar

## Introducción

### 1.1 PROBLEMAS ANALÍTICOS

Un químico analítico puede enfrentarse a dos tipos de problemas: en ocasiones se le pide sólo una respuesta cualitativa; por ejemplo, la presencia de boro en agua destilada es muy perjudicial en la producción de componentes microelectrónicos (“¿Contiene boro esta muestra de agua destilada?”). Por otro lado un problema común en la ciencia forense es la comparación de muestras de suelo (“¿Estas dos muestras de suelo pueden proceder del mismo lugar?”). En otros casos, los problemas que se le plantean son cuantitativos (“¿Cuánta albúmina hay en esta muestra de suero sanguíneo?”, “¿cuánto plomo hay en esta muestra de agua del grifo?”, “esta muestra de acero contiene pequeñas cantidades de cromo, tungsteno y manganeso; ¿cuánto de cada uno?”). Éstos son ejemplos típicos de análisis cuantitativos de uno o varios componentes.

La química analítica moderna es una ciencia predominantemente **cuantitativa**. Es obvio que en muchos casos una respuesta cuantitativa será mucho más valiosa que una cualitativa; para un analista puede ser útil decir que ha detectado boro en una muestra de agua destilada, pero es mucho más útil para él poder definir *cuánto* boro se encuentra presente en dicha muestra. La persona que solicita el análisis podría, una vez que cuenta con esta respuesta cuantitativa, juzgar si la concentración de boro es preocupante, considerar cómo se puede reducir, etc.; pero si sólo supiera que hay *algo* de boro le sería difícil estimar la significación del resultado. En otros casos sólo tiene valor un resultado cuantitativo. Por ejemplo, prácticamente todas las muestras de suero sanguíneo humano contienen albúmina; la pregunta es, ¿cuánta?

Es importante observar que aun cuando se requiera una respuesta cualitativa, se utilizan métodos cuantitativos para obtenerla. Esto se ilustra en los ejemplos expuestos al principio de esta sección. En realidad, un analista nunca diría simplemente “Yo puedo o no puedo detectar boro en esta muestra de agua”. Utilizaría un método cuantitativo capaz de detectar boro a niveles, supongamos, de  $1 \mu\text{g/ml}$ . Si la prueba diese un resultado negativo, podría describirse en la forma “Esta muestra contiene menos de  $1 \mu\text{g/ml}$  de boro”. Si el ensayo dio un resultado positivo, se podría decir que la muestra contiene al menos  $1 \mu\text{g/ml}$  de boro

Para comparar dos muestras de suelo se podría utilizar aproximaciones cuantitativas mucho más complejas. Por ejemplo, las muestras podrían estar sujetas a un análisis de tamaños de partículas que determinaría las proporciones de partículas de suelo clasificadas en cierto número de tamaños, digamos diez. Entonces cada muestra se caracterizaría por estos diez datos. Se puede usar procedimientos muy complejos (véase el Cap. 7) para brindar una valuación cuantitativa de su similitud, y se puede obtener una estimación de la probabilidad de que las muestras tengan un origen común, si se dispone de datos comparables de muestras de suelo que *se sabe* que son compatibles.

## 1.2 ERRORES EN EL ANÁLISIS CUANTITATIVO

Una vez que aceptamos que los análisis cuantitativos jugarán un papel predominante en cualquier laboratorio analítico, debemos aceptar también que los errores que aparezcan en tales estudios son de gran importancia. Nuestro principio guía será que *no existen resultados cuantitativos válidos si no van acompañados de alguna estimación de los errores inherentes a ellos*. Desde luego, este principio se aplica no solo a la química analítica, sino a cualquier campo de estudio en el que se obtengan resultados numéricos experimentales. Podemos examinar rápidamente una serie de ejemplos sencillos que ilustran el principio y que además plantean los tipos de problemas estadísticos que encontraremos y resolveremos en los próximos capítulos.

Supongamos que un químico sintetiza un reactivo analítico creyendo que es completamente nuevo; lo estudia utilizando un método espectrométrico y el compuesto da un valor de 104 (normalmente, la mayoría de nuestros resultados se citarán en unidades elegidas con todo cuidado, pero en este ejemplo hipotético pueden utilizarse unidades arbitrarias). Al revisar los libros de referencia, el químico encuentra que ningún compuesto descubierto hasta el momento ha arrojado un valor de más de 100 cuando se ha estudiado con el mismo método en las mismas condiciones experimentales. La pregunta que surge naturalmente es: ¿nuestro químico descubrió en realidad un compuesto nuevo? Por supuesto, la respuesta a esta pregunta reside en el grado de confianza que podemos asignar al valor experimental de 104. ¿Qué errores están asociados con él? Si el estudio posterior indica que el resultado es correcto dentro de 2 unidades (arbitrarias), es decir, que el verdadero valor se encuentra en el intervalo  $104 \pm 2$ , entonces es posible que se haya caracterizado un nuevo material. Pero si las investigaciones muestran que el error puede alcanzar hasta 10 unidades (es decir,  $104 \pm 10$ ), entonces es muy probable que el verdadero valor sea en realidad menor que 100, en cuyo caso el nuevo descubrimiento dista mucho de ser cierto. En otras palabras, es esencial conocer los errores experimentales (en este caso como en cualquier otro) para interpretar adecuadamente los resultados. En términos estadísticos, este ejemplo conlleva la comparación de los resultados experimentales con un valor supuesto o de referencia: este tema se estudia con detalle en el capítulo 3.

Una situación más común es la del analista que lleva a cabo varias determinaciones repetidas en el curso de un solo análisis. (El valor y la significación de tales repeticiones se verá con detalle en el próximo capítulo.) Suponga que un analista efectúa cuatro veces un experimento volumétrico y obtiene valores de 24.69, 24.73, 24.77 y 25.39 ml. El primer aspecto por señalar es que los valores de titulación se redondean a 0.01 ml; este aspecto también se analiza en el capítulo 2. También es evidente de inmediato que los cuatro valores son distintos debido a los errores inherentes a las medidas, y que el cuarto valor (25.39 ml)

es sustancialmente diferente de los otros tres. Aquí surge la pregunta: ¿puede rechazarse sin riesgo este cuarto valor de manera que, por ejemplo, el valor medio se presente como 24.73 ml? En términos estadísticos, ¿el valor de 25.39 ml es un resultado anómalo? El importante tema del rechazo de los resultados anómalos se analiza con detalle en el capítulo 3.

Otro problema frecuente es la comparación de dos o más conjuntos de resultados. Suponga que un analista mide el contenido de vanadio en una muestra de acero usando dos métodos distintos. Con el primero obtiene un valor medio de 1.04%, con un error estimado del 0.07%; con el segundo obtiene un valor medio del 0.95% y un error del 0.04%. Al comparar estos resultados surgen varias preguntas: ¿los dos valores medios son significativamente diferentes, o son indistinguibles dentro de los límites de error experimental? ¿Uno de los métodos es bastante menos propenso a errores que el otro? ¿Cuál de los dos valores medios está en realidad más cerca de la verdad? De nuevo, en el capítulo 3 se estudian éstas y otras preguntas relacionadas.

Para concluir esta sección notemos que muchos análisis se basan en métodos gráficos. En lugar de medir repetidamente la misma muestra, realizamos una serie de mediciones sobre un pequeño grupo de estándares los cuales tienen concentraciones conocidas que cubren un intervalo considerable; de esta manera establecemos una curva de calibración que puede utilizarse para estimar la concentración de muestras estudiadas con el mismo procedimiento. En la práctica, por supuesto, todas las mediciones (las que se basan en estándares y las que se efectúan en muestras) estarán sujetas a errores. Por ejemplo, es necesario calcular los errores en que se incurre al trazar la curva de calibración, estimar el error de la concentración de una muestra determinada utilizando la curva, y estimar el límite de detección del método, es decir, la cantidad de analito más pequeña que puede detectarse con un grado de confianza concreto. Estos procedimientos, especialmente frecuentes en análisis instrumental, se describen en el capítulo 5.

Estos ejemplos representan sólo una parte de los problemas que pueden surgir de la ocurrencia de errores experimentales en el análisis cuantitativo. No obstante, como hemos visto, los problemas deben resolverse si deseamos que los datos cuantitativos tengan algún significado real. Por lo tanto, resulta evidente que debemos estudiar con más detalle los diferentes tipos de errores.

## 1.3 TIPOS DE ERRORES

Los científicos experimentales hacen una distinción fundamental entre tres tipos de errores: **crasos**, **aleatorios** y **sistemáticos**. Los errores crasos se describen con facilidad; pueden ocurrirse como errores tan graves que no queda otra alternativa más que abandonar el experimento y empezar de nuevo. Como ejemplos se podrían incluir la avería total de un instrumento, la caída o el derramamiento accidental de una muestra muy importante, o descubrir durante el desarrollo de un experimento que un reactivo que se suponía puro, en realidad estaba contaminado. Tales errores (¡que ocurren ocasionalmente incluso en los laboratorios mejor controlados!) normalmente se reconocen con mucha facilidad. En consecuencia, en nuestro análisis sólo tenemos que distinguir con detenimiento entre los errores **aleatorios** y los **sistemáticos**.

Podemos hacer mejor esta distinción mediante el estudio cuidadoso de una situación experimental real. Cuatro estudiantes (A-D) realizan un análisis en el que titulan



exactamente 10.00 ml de hidróxido de sodio exactamente 0.1M\* con ácido clorhídrico exactamente 0.1M. Cada uno realiza cinco réplicas del análisis y obtiene los resultados que se muestran en la tabla 1.1.

Los resultados obtenidos por el estudiante A tienen dos características importantes. Primero, todos están muy próximos; todos caen entre 10.08 y 10.12 ml. En términos generales, diríamos que los resultados son altamente *reproducibles*. La segunda característica distintiva es que *todos son demasiado altos*; en este experimento (algo inusual) conocemos de antemano la respuesta correcta: 10.00 ml. Resulta evidente que han surgido dos tipos de errores completamente distintos en el experimento de este estudiante. En primer lugar existen **errores aleatorios**, los cuales provocan que los resultados individuales caigan a ambos lados del valor medio (en este caso 10.10 ml). Los estadísticos afirman que los errores aleatorios afectan la **precisión**, o **reproducibilidad** de un experimento. En el caso del estudiante A, queda claro que los errores aleatorios son pequeños y, por lo tanto, decimos que los resultados son **precisos**. Sin embargo, también existen **errores sistemáticos**, los cuales provocan que todos los resultados sean erróneos en el mismo sentido (en este caso todos son demasiado altos). Los errores sistemáticos afectan la **exactitud**, es decir, la *proximidad al valor verdadero*. En muchos experimentos, los errores aleatorios y sistemáticos no se detectan fácilmente con sólo observar los resultados, sino que también tienen orígenes muy distintos en cuanto a la técnica experimental y al equipo que se utiliza. Antes de examinar las causas de los errores en este experimento, podemos analizar brevemente los resultados obtenidos por los estudiantes B a D. El estudiante B ha obtenido resultados que se encuentran en contraposición directa con los del estudiante A. El promedio de los cinco resultados (10.01 ml) está muy próximo al valor verdadero, de manera que podemos caracterizar los datos como exactos, sin errores sistemáticos sustanciales. Sin embargo, la variedad de los resultados es muy grande, lo que indica una precisión insatisfactoria y la presencia de errores aleatorios sustanciales. La comparación de estos resultados con los obtenidos por el estudiante A muestra con claridad que los errores aleatorios y los sistemáticos pueden ocurrir independientemente uno del otro. Esta conclusión se refuerza con los datos de los estudiantes C y D. El trabajo del estudiante C no es ni preciso (intervalo de 9.69 a 10.19 ml) ni exacto (media 9.90 ml). El estudiante D ha logrado a la vez resultados precisos (intervalo de 9.97 a 10.04 ml) y exactos (media 10.01 ml). La distinción entre errores aleatorios y sistemáticos, y entre precisión y exactitud, se resume en la figura 1.1.

El experimento aparentemente sencillo, descrito en los párrafos anteriores, merece varios comentarios adicionales. Es muy importante mencionar que las palabras **precisión** y **exactitud** tienen significados completamente diferentes en la teoría de errores, mientras que se usan de manera indistinta en el lenguaje cotidiano. Por ejemplo, en varios diccionarios se menciona la exactitud como una definición de precisión, y viceversa. Esta desafortunada confusión hace más difícil recordar la diferencia principal entre los errores sistemáticos y los aleatorios, a los cuales se refieren las palabras exactitud y precisión, respectivamente. Mencionaremos dos problemas adicionales de nomenclatura. En algunos textos se habla de errores “determinados” e “indeterminados” con el mismo significado que errores sistemáticos y aleatorios, respectivamente. Así mismo, aunque utilizamos con anterioridad la palabra “reproducibilidad” como una definición aproximada de la precisión, el convenio actual hace una distinción ciudadana entre **reproducibilidad** y **repetitividad**.

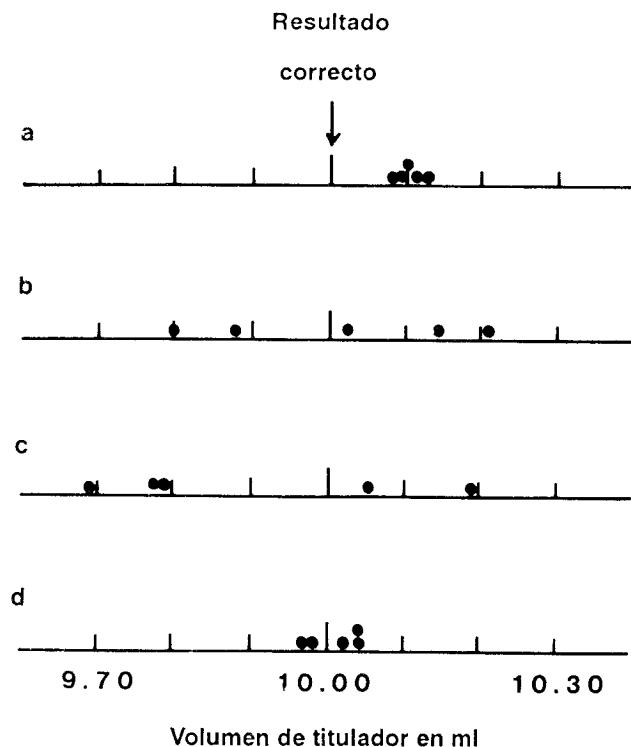
\* El símbolo 1M indica 1 mol del material especificado por litro de solución.

Tabla 1.1 Errores sistemáticos y aleatorios

Estudiante	Resultados (ml)	Comentario
A	10.08	Preciso pero inexacto
	10.11	
	10.09	
	10.10	
	10.12	
B	9.88	Exacto pero impreciso
	10.14	
	10.02	
	9.80	
	10.21	
C	10.19	Inexacto e impreciso
	9.79	
	9.69	
	10.05	
	9.78	
D	10.04	Exacto y preciso
	9.98	
	10.02	
	9.97	
	10.04	

Podemos aclarar esta diferencia con una extensión del experimento anterior. Por el camino normal el estudiante A, por ejemplo realizaría las cinco medidas repetidas en una sucesión rápida; pero es probable que no tardara más de una hora en realizar el ejercicio completo. Utilizaría la misma serie de disoluciones y el mismo material de vidrio a lo largo del experimento; añadiría la misma preparación del indicador a cada matraz de titulación, y permanecerían iguales la temperatura, la humedad y demás condiciones de laboratorio. En tales circunstancias, la precisión medida sería la precisión *dentro de rachas*; esto se denomina **repetitividad**. Sin embargo, suponga que por alguna razón el mismo estudiante realizó las titulaciones en cinco ocasiones diferentes. Es probable que en tales circunstancias usara recipientes de vidrio y preparaciones de indicador diferentes; así mismo, las condiciones del laboratorio también pudieron cambiar de una ocasión a otra. Por eso en este caso no sería sorprendente encontrar una gran variabilidad en los resultados. Este conjunto de datos reflejaría la precisión *entre rachas* del método, y esto es a lo que se debería asignar el término **reproducibilidad**. En la tabla 1.2 se resumen las definiciones de los términos utilizados y las relaciones entre ellos.

Se debe aprender otra lección del experimento de titulación. Es fácil apreciar que los valores obtenidos por el estudiante C son inaceptables, y que los del estudiante D son los más aceptables. Sin embargo, en ocasiones puede ocurrir que haya dos métodos disponibles para un análisis concreto, de los cuales se piense que uno sea preciso pero inexacto, mientras que el otro, exacto pero impreciso. En otras palabras, es posible que tengamos que elegir entre los tipos de los resultados obtenidos por los estudiantes A y B respectivamente. Entonces, ¿qué tipo de resultado es preferible? No es posible dar una respuesta categórica a esta pregunta, aunque sólo sea porque la elección del método analítico se basará, en la práctica, en el costo, la facilidad de automatización, la rapidez del



**Figura 1.1** Exactitud y precisión; representación gráfica de los datos de la tabla 1.1. En (a) los datos son precisos pero inexactos; en (b) exactos pero imprecisos; en (c) inexactos e imprecisos, y en (d) exactos y precisos.

análisis y otros factores que se encuentran fuera de nuestra evaluación elemental. Sin embargo, es importante darse cuenta de que un método que se encuentre sustancialmente libre de errores sistemáticos puede aún, si es muy impreciso, proporcionar un valor medio que esté (por azar) bastante alejado del valor correcto. Por otra parte, un método que es preciso, pero inexacto (estudiante A) puede convertirse en uno que sea preciso y exacto (estudiante D), si se descubren y eliminan los errores sistemáticos. También existirán casos en donde, debido a que las medidas que se obtienen son absolutamente nuevas, no sea factible controlar alguno de los errores sistemáticos. Los errores aleatorios nunca se pueden eliminar, aunque mediante una técnica cuidadosa se pueden minimizar, y realizando medidas repetidas se puede medir y evaluar su importancia. En muchos casos pueden eliminarse los errores sistemáticos mediante controles adecuados de nuestra técnica expe-

**Tabla 1.2** Resumen de definiciones

Tipos de errores	
Aleatorios	Sistemáticos
Afectan la precisión	Afectan la exactitud
La precisión dentro de la racha es la <b>repetitividad</b>	Proximidad a la verdad
La precisión entre las rachas es la <b>reproducibilidad</b>	
También se conocen como errores indeterminados	También se conocen como errores determinados

rimental y de nuestro equipo. En la próxima sección se analiza esta importante distinción adicional entre los dos grandes tipos de errores.

## 1.4 ERRORES SISTEMÁTICOS Y ALEATORIOS EN EL ANÁLISIS VOLUMÉTRICO

Mediante el ejemplo del experimento volumétrico de los estudiantes se muestra con claridad que los errores sistemáticos y aleatorios pueden ocurrir independientemente unos de otros y surgir en diferentes etapas del experimento. Ya que la volumetría es un procedimiento relativamente simple y todavía ampliamente usado, es importante examinarlo con detalle en este contexto. Se puede considerar que un análisis volumétrico es completo cuando involucra los siguientes pasos (algunos de los cuales se han omitido en el ejemplo de la sección precedente con el fin de hacerlo más sencillo).

(1) Preparación de una solución estándar de uno de los reactivos. Esto supone (a) pesar un pesasustancias o recipiente similar que contenga algún material sólido; (b) transferir el material sólido a un matraz aforado y pesar de nuevo el pesasustancias con el fin de obtener por diferencia el peso del sólido transferido, y (c) llenar el matraz con agua destilada hasta el enrase (suponiendo que se contempla una titulación acuosa).

(2) Transferir una alícuota del material estándar a un matraz de titulación con ayuda de una pipeta. Esto supone (a) llenar la pipeta hasta el enrase adecuado, y (b) vaciar el contenido de una manera específica en el matraz de titulación.

(3) Titulación del líquido en el matraz con una solución del otro reactivo añadido de una bureta. Esto implica (a) llenar la bureta y permitir que el líquido contenido en ella escurra hasta que el menisco alcance un nivel constante; (b) añadir unas cuantas gotas de solución de indicador al matraz de titulación; (c) leer el volumen inicial de la bureta; (d) añadir poco a poco el líquido de la bureta al matraz de titulación hasta que se juzgue que se ha alcanzado el punto final, y (e) medir el nivel final del líquido en la bureta.

Si bien un análisis elemental de este tipo implica diez pasos diferentes, por lo general los siete últimos se repiten, como hemos visto, varias veces. En principio, podríamos examinar cada paso para evaluar los errores aleatorios y sistemáticos que pudieran ocurrir. En la práctica, es más sencillo examinar por separado las etapas que utilizan pesadas [pasos 1 (a) y 1 (b)], y las etapas restantes que conllevan el uso de equipo volumétrico. El lector advertirá que la intención no es proporcionar descripciones detalladas de las técnicas experimentales utilizadas en las diferentes etapas. De igual manera, no se exponen métodos para calibrar pesas, material de vidrio, etc. (Estos temas se tratan con detalle en los libros citados en la bibliografía al final del capítulo.) Entre las contribuciones para detectar los errores, son de importancia las tolerancias de las pesas utilizadas en los pasos gravimétricos, y del material de vidrio volumétrico. Algunas instituciones tales como la *British Standards Institution* (BSI) y la *American Society for Testing and Materials* (ASTM) han publicado especificaciones estándar para estas tolerancias. Algunos valores de tolerancia típica se exponen en la tabla 1.3. La tolerancia de una pesa de 100 g de alta calidad puede ser tan pequeña como  $\pm 0.25$  mg, aunque para un peso usado en tareas de rutina la tolerancia podría ser hasta cuatro veces más grande. De igual manera, la tolerancia para un matraz aforado de grado A de 250 ml es  $\pm 0.12$  ml; el material volumétrico de grado B por lo regular, presenta tolerancias dos veces más grandes que el material de vidrio grado A. Si una pesa o una pieza de material de vidrio está dentro de los límites de tolerancia, pero no posee

**Tabla 1.3** Ejemplos de tolerancias de pesos y de material de vidrio

Artículo	Tolerancia
Pesos (Clase 1)	$\pm 0.25$ mg
100 g	0.034 mg
1 g	0.010 mg
10 mg	
Pesos (Clase 3)	$\pm 1.0$ mg
100 g	0.10 mg
1 g	0.030 mg
10 mg	
Material de vidrio (Grado A)	$\pm 0.25$ mg
Probeta graduada de 50 ml	0.12 mg
Matraz aforado de 250 ml	0.03 mg
Pipeta de 25 ml	0.05 mg
Bureta de 50 ml	

(Datos obtenidos de las Especificaciones estándar de ASTM E617-78 y E694-79; véase la bibliografía.)

exactamente el peso o volumen correcto, surgirá un error sistemático. Así pues, si el matraz aforado tiene en realidad un volumen de 249.95 ml, este error se reflejará en el resultado de todos los experimentos basados en el uso de dicho matraz. La repetición del experimento no revelará el error, ya que en cada repetición se supondrá que el volumen es de 250.00 ml, cuando de hecho es menor que éste. No obstante, si se comparan los resultados de un experimento en el que se ha utilizado este matraz con los de otros experimentos (p. ej., en otros laboratorios) realizados con otros matraces, entonces el hecho de que todos los matraces tengan volúmenes ligeramente diferentes contribuirá a la variación aleatoria como ya se mencionó, los errores aleatorios entre rachas afectan la reproducibilidad de los resultados.

Por lo regular los procedimientos de pesada se encuentran asociados con errores **aleatorios** muy pequeños. Es muy común que en tareas de laboratorio rutinarias se utilicen balanzas de cuatro cifras, y el error aleatorio que se comete no debería ser mayor que  $\pm 0.0001$ – $0.0002$  g (en el próximo capítulo se describe con detalle el término estadístico empleado para expresar errores aleatorios). Debido a que la cantidad que se pesa por lo regular se encuentra alrededor de 1 g o más, resulta evidente que el error aleatorio, expresado en un porcentaje de la pesada efectuada, no es mayor del 0.02%. Un buen material estándar para el análisis volumétrico debería tener (entre otras características) un peso fórmula tan alto como sea posible, de manera que se minimicen estos errores aleatorios de pesada cuando se prepare una solución de una molaridad dada. En algunos análisis se utilizan “microbalanzas” para pesar cantidades de pocos miligramos; pero los errores de pesada cometidos son sólo unos pocos microgramos.

Los errores **sistemáticos** en las pesadas pueden ser apreciables y proceder de una serie de fuentes establecidas, las cuales incluyen la adsorción de la humedad en la superficie del recipiente de pesada; los errores provocados al no permitir que los recipientes calentados se enfrien a la misma temperatura que la balanza antes de pesar (este error es muy común en gravimetría cuando se pesan crisoles); las pesas oxidadas o contaminadas por el polvo; y el efecto boyante de la atmósfera, que actúa en diferente medida sobre objetos de distinta densidad. Con el fin de lograr un trabajo más exacto, se debe calibrar las pesas en relación

con estándares suministrados por cuerpos estatutarios y autoridades de normativas (véanse los párrafos anteriores). Esta calibración puede ser muy exacta, por ejemplo, de  $\pm 0.01$  mg para pesadas en el intervalo de 1 a 10 g. El efecto atmosférico puede ser sustancial; por ejemplo, Skoog y West (véase la bibliografía) presentan una muestra de un líquido orgánico de densidad de 0.92 g/ml que pesa 1.2100 g al aire, y que pesaría 1.2114 g al vacío, una diferencia de más del 0.1%. Aparte de usar procedimientos de calibración, los cuales se analizarán en la próxima sección, pueden tomarse algunas precauciones experimentales sencillas para minimizar estos errores sistemáticos; la más importante es realizar todas las pesadas **por diferencia**, es decir, se pesa primero un pesafiltros con la muestra contenida en él, y después sin ella. La **diferencia** entre las dos pesadas proporciona el peso de la muestra, y también elimina los errores sistemáticos que surjan, por ejemplo de la humedad u otros contaminantes en la superficie del recipiente (véase también la Sec. 2.8). *Si se toman estas precauciones*, los errores producidos en los pasos de la pesada serán mínimos, y es probable que en muchos experimentos volumétricos estos errores sean despreciables comparados con los que surjan por el uso del equipo volumétrico. De hecho, los métodos gravimétricos se usan generalmente para calibrar el material de vidrio volumétrico, pesando (en condiciones estándar) el agua que contenga; resulta evidente que este procedimiento es válido tan sólo debido a que los errores gravimétricos son despreciables comparados con los volumétricos.

En los procedimientos volumétricos los errores aleatorios provienen del uso de material de vidrio. Al llenar un matraz aforado de 250 ml hasta la marca del aforo, el error (es decir, la distancia entre el menisco y la marca) puede ser aproximadamente de  $\pm 0.03$  cm en el cuello del matraz de diámetro de ca. 1.5 cm. Éste corresponde a un error en volumen de 0.05 ml, sólo un 0.02% del volumen total del matraz. De igual manera, al llenar una pipeta de transferencia de 25 ml el error aleatorio no debería exceder de 0.03 cm en una caña de 0.5 cm de diámetro; ésta proporciona un error en volumen de ca. 0.006 ml, el 0.024% del volumen total. El error en la lectura de una bureta (graduada en divisiones de 0.1 ml) es quizás de 0.01 a 0.02 ml. Cada titulación implica dos lecturas (los errores *no* son simplemente aditivos, véase el capítulo 2); si el volumen de titulación es aproximadamente 25 ml, el porcentaje de error es nuevamente muy pequeño. Es evidente que las condiciones experimentales deberían planificarse de manera que el volumen de titulador no sea demasiado pequeño (no menos de 10 ml), ya que de otra manera los errores serían considerables. (Esta precaución es análoga a la de utilizar un compuesto estándar de alto peso molecular para minimizar el error de pesada.) Además, aunque un análisis volumétrico suponga varios pasos, y en cada uno de ellos se use una pieza de material de vidrio volumétrico, es obvio que los errores aleatorios deberían ser pequeños si los experimentos se realizan con cuidado. En la práctica, un buen análisis volumétrico debería tener una desviación estándar relativa (véase el capítulo 2) de no más de 0.1%. (Hasta hace poco, tal precisión normalmente no se alcanzaba con los métodos de análisis instrumental, y aún no es común, y constituyó una de las mayores ventajas de las técnicas analíticas “clásicas”.) Debemos recordar que, con habilidad y tomando todas las precauciones, los métodos clásicos pueden dar resultados con desviaciones estándar relativas, tan bajas como del 0.01%.

Los procedimientos volumétricos incorporan varias fuentes importantes de error sistemático. Entre los principales se encuentran los errores de vaciado en el uso de material de vidrio volumétrico; los errores de calibración en el material de vidrio, y los “errores del indicador”. Quizás el error más común en el análisis volumétrico de rutina sea no dejar pasar el tiempo suficiente para vaciar una pipeta, o para estabilizar el nivel del menisco en

la bureta. También es importante recordar que hay dos tipos de pipetas, las que se vacían por evacuación, y las pipetas de soplado, en las que el último líquido remanente debe ser expulsado por la fuerza. Si se confunden los dos tipos, por ejemplo, soplando en una pipeta de vaciado, ¡se incurriría ciertamente en un error craso! Los errores por vaciado tienen un efecto sistemático y otro aleatorio: el volumen liberado invariablemente es menor que el que debería ser. La temperatura a la que se realiza un experimento tiene dos efectos. El equipo volumétrico se calibra en forma convencional a 20°C, pero la temperatura en un laboratorio analítico puede diferir con facilidad de ésta en varios grados, y muchos experimentos, por ejemplo los análisis bioquímicos, se llevan a cabo en “habitaciones frías” a ca. de 4°C. En segundo lugar, la temperatura afecta tanto al volumen de material de vidrio como a la densidad de los líquidos. El coeficiente de expansión para soluciones acuosas diluidas es aproximadamente del 0.025% por grado, mientras que para recipientes de vidrio sódico variará casi un 0.003% por grado en volumen, y para recipientes de vidrio borosilicatado la variación será del 0.001% por grado. Resulta evidente que los cambios en los volúmenes del material de vidrio serán importantes sólo en trabajos de gran exactitud, e incluso sólo si la temperatura es muy diferente de 20°C. Además, los efectos de la expansión de las soluciones se autocompensarán en gran parte si todas las soluciones se mantienen a la misma temperatura. El efecto es más obvio en soluciones no acuosas.

Los errores de indicador pueden ser muy importantes: quizás más grandes que los errores aleatorios en un análisis volumétrico típico. Por ejemplo, en la titulación de ácido clorhídrico al 0.1M, con hidróxido sódico al 0.1M esperamos que el punto final corresponda con un pH de 7. Sin embargo, en la práctica, estimamos este punto final usando un indicador como el naranja de metilo. En experimentos diferentes se muestra que esta sustancia cambia de color sobre un intervalo de pH ca. 3-4. Por lo tanto, si la titulación se realiza añadiendo la base al ácido, el indicador conducirá a un punto final aparente cuando el pH sea ca. 3.5, es decir, justo antes del verdadero punto final. Es probable que el error sistemático que se ocasiona aquí sea del 0.2%. Si la titulación se realiza a la inversa, es decir añadiendo ácido a la base, el punto final indicado por el naranja de metilo estará en realidad un poco más alejado del punto final verdadero, debido a la necesidad de alcanzar un pH ácido para conseguir el cambio de color. Así pues, el error sistemático debido al indicador puede ser positivo o negativo de acuerdo con el diseño del experimento; para un procedimiento dado será siempre positivo o siempre negativo. En cualquier caso, se puede evaluar y corregir el error realizando un experimento en **blanco**, es decir, determinado qué cantidad de base (o ácido) es necesaria para producir el cambio de color del indicador ante la *ausencia* del ácido (base). Un error que se menciona rara vez es el de la “última gota”. No se sabe con certeza *qué cantidad* de la última gota se requiere para alcanzar el punto final. Ya que se necesitará un volumen mayor o menor al de la mitad de la última gota, puede reducirse este error restando la mitad del volumen de una gota.

En cualquier procedimiento analítico, ya sea clásico o instrumental, es posible considerar y estimar la fuentes de error sistemático y aleatorio que surjan en cada una de las etapas del experimento; es deseable que el analista haga esto con un cuidadoso diseño experimental, ya que le permitirá evitar fuentes de error mayores (véanse las Secs. 1.5 y 1.6). Sin embargo, vale la pena señalar que en los análisis volumétricos los errores son bastante inusuales en el sentido de que no implican ninguna etapa individual que tenga un error mucho mayor que los errores de otras etapas. En muchos otros análisis, en la práctica el error global está dominado por el error generado en un único paso; este aspecto se analizará en el próximo capítulo.

## 1.5 EL MANEJO DE ERRORES SISTEMÁTICOS

Gran parte del contenido de este libro tratará de la evaluación de errores aleatorios, los cuales pueden estudiarse mediante un amplio conjunto de métodos estadísticos. En muchos casos supone, por conveniencia, que están ausentes los errores sistemáticos (aunque se describirán los métodos que contrastan la existencia de errores sistemáticos). Es necesario analizar en este momento los errores sistemáticos con más detalle: cómo surgen, y cómo pueden abordarse. En el ejemplo del análisis volumétrico expuesto en la sección 1.3 se muestra con claridad que los errores sistemáticos hacen que el valor medio de una serie de mediciones repetidas se desvíe del verdadero valor. Se deduce que: (a) a diferencia de los errores aleatorios, los sistemáticos no se pueden apreciar con la simple repetición de mediciones, y (b) a menos que se conozca de antemano el resultado verdadero de un análisis (una situación muy improbable!), pueden existir errores sistemáticos muy grandes, que pasen inadvertidos si no se toman las debidas precauciones. En otras palabras, es demasiado fácil pasar por alto las fuentes principales de error sistemático. Un pequeño conjunto de ejemplos clarificarán tanto los posibles problemas y sus soluciones.

En los últimos años se ha mostrado mucho interés en las concentraciones de metales de transición en muestras biológicas tales como el suero sanguíneo. Por ejemplo, se han realizado muchas determinaciones de los niveles de cromo en el suero con resultados sorprendentes. Diferentes químicos han obtenido concentraciones de cromo, que varían desde < 1 a ca. de 200 ng/ml, todos ellos estudiando muestras de suero combinadas procedentes de individuos sanos. En general, se obtuvieron los resultados más bajos recientemente y se hizo evidente que los resultados más altos, al menos en parte, se debían a la contaminación de las muestras por el cromo procedente de las jeringas de acero inoxidable, los tapones de los tubos, etc. La determinación de trazas de cromo, por ejemplo, por espectrometría de absorción atómica es, en principio, relativamente directa, y no hay duda en que los resultados alcanzados por cada grupo de químicos son satisfactorios en términos de precisión, aunque en varios casos se pasó por alto el gran error sistemático introducido por la contaminación. Los errores metodológicos sistemáticos de este tipo son comunes: el lavado incompleto de un precipitado en un análisis gravimétrico, y el error producido por el indicador en un análisis volumétrico (véase la Sec. 1.4) son ejemplos muy conocidos.

A menudo se presenta otra clase de errores sistemáticos, que surgen cuando se realizan hipótesis falsas sobre la exactitud de un instrumento analítico. Los analistas experimentados saben demasiado bien que los monocromadores en los espectrómetros se desajustan paulatinamente de manera que son frecuentes los errores de varios nanómetros en las longitudes de onda seleccionadas, aún así se realizan muchos análisis fotométricos sin efectuar los controles adecuados. Los aparatos muy sencillos tales como material de vidrio volumétrico, cronómetros, pH-metros y termómetros, pueden mostrar errores sistemáticos sustanciales, si bien una gran mayoría del personal de laboratorio utiliza con regularidad estos instrumentos como si fueran siempre exactos. Además, la disponibilidad creciente de instrumentos controlados por microprocesadores o microcomputadores ha reducido al mínimo el número de operaciones y el nivel de pericia de los operadores. En estas circunstancias, es irresistible la tentación de considerar que los resultados de los instrumentos son perfectos, a pesar de que de todas maneras están sujetos a errores sistemáticos (a menos que posean suficiente “inteligencia” para poder autocalibrarse, véase la sección 1.7).

Los errores sistemáticos no sólo surgen de procedimientos y de aparatos; pueden surgir también de sesgos humanos. Algunos químicos padecen algún grado de astigmatismo o daltonismo (lo último es más común en los hombres que en las mujeres) que podría provocar errores en la lectura de instrumentos y otras observaciones. Varios autores han registrado numerosos prejuicios, por ejemplo, una tendencia que favorezca los números pares sobre los impares, o incluso los dígitos 0 y 5 sobre otros dígitos, en la presentación de resultados. Por lo tanto, queda claro que las diversas clases de errores sistemáticos son un riesgo constante, y a menudo oculto, para el analista, de manera que para minimizarlos deben examinarse los sucesivos pasos con sumo cuidado. Se dispone de diferentes aproximaciones a este problema, y se debe considerar cualquiera de ellas o todas en cada uno de los procedimientos analíticos.

Deberían tomarse las primeras precauciones antes de empezar cualquier trabajo experimental. El analista debería considerar con cuidado cualquier etapa del experimento que esté a punto de realizar, el aparato que se utilice, el muestreo y el procedimiento analítico que se vaya a adoptar. En esta primera etapa, debería intentar identificar las fuentes probables de error sistemático, tales como las funciones del instrumento que necesitan calibración, y los pasos del procedimiento analítico donde los errores tienen mayor probabilidad de ocurrir. También debe considerar las pruebas que puedan efectuarse para verificar los errores sistemáticos. Pueden ser de gran valor las previsiones de este tipo (veremos en la próxima sección que debe prestarse igual atención a las fuentes de error aleatorio) y, por lo regular, vale la pena el tiempo invertido. Por ejemplo, una buena reflexión de este tipo podría haber revelado la posibilidad de que existiera contaminación en el experimento de determinación de cromo descrito con anterioridad.

La segunda acción contra los errores sistemáticos radica en el diseño del experimento en cada etapa. Ya se ha visto (§ec. 1.4) que la pesada por diferencia puede eliminar algunos errores sistemáticos gravimétricos: puede suponerse que estos errores existen con la misma extensión en ambas pesadas, de manera que el proceso de sustracción los elimina. Otro ejemplo de planificación experimental lo proporciona el error de la longitud de onda del espectrómetro descrito anteriormente. Existen dos procedimientos para determinar la concentración de una muestra por espectrometría de absorción. En el primero se estudia la muestra en una célula del espectrómetro de 1 cm a una longitud de onda de 400 nm, y la concentración del componente ensayado se determina a partir de la ecuación conocida  $A = \epsilon cl$  [donde  $A$ ,  $\epsilon$ ,  $c$  y  $l$  son respectivamente la absorbancia medida, el valor de referencia aceptado de la absorbancia molar (unidades  $\text{l.mol}^{-1}.\text{cm}^{-1}$ ) del componente ensayado, la concentración molar de este analito, y el paso (cm) de la célula del espectrómetro]. Aquí pueden surgir varios errores sistemáticos; la longitud de onda podría ser, como ya se ha comentado, por ejemplo, de 405 nm en vez de 400 nm, con lo cual el valor de referencia de  $\epsilon$  es inadecuado; el valor de  $\epsilon$  podría ser incorrecto en cualquier caso; el paso óptico de la célula podría no ser exactamente de 1 cm. De manera alternativa, el analista puede emplear una serie de diluciones de la sustancia ensayada de concentración conocida, y medir la absorbancia de cada una a 400 nm. (Una de estas soluciones de calibración sería un blanco, es decir, contendría todos los reactivos analíticos excepto el de la sustancia que se ensaya; véase la sección 1.4.) Entonces, se podría utilizar los resultados con el fin de construir una línea de calibración para el análisis de la muestra ensayada en las mismas condiciones experimentales. Esta importante aproximación al análisis instrumental se describe con detalle en el capítulo 5. Cuando se utiliza este segundo método, no se requiere el valor de  $\epsilon$  y se anulan los errores debidos a los desplazamientos de las longitudes de

onda, errores de absorbancia e inexactitud del paso óptico, ya que ocurren igualmente en la calibración y en la medición de la muestra. Todas las fuentes principales de error sistemático son, sólo en principio, eliminadas con tal de que las condiciones para la muestra y la calibración sean de hecho equivalentes (por ejemplo, suponemos que la longitud de onda y las escalas de absorbancia no se alteran *durante el experimento*.)

Por último, quizás la protección más efectiva contra los errores sistemáticos sea el uso de materiales de referencia y métodos estándar. Antes de comenzar el experimento, se calibra cada pieza de equipo mediante un proceso adecuado. Ya hemos visto cómo puede calibrarse el equipo volumétrico mediante el uso de métodos gravimétricos. Así mismo, se puede calibrar la escala de las longitudes de onda del espectrómetro con la ayuda de fuentes de luz estándar, las cuales muestran líneas de emisión estrechas a longitudes de onda bien establecidas, y la escala de absorbancia del espectrómetro se calibra con filtros estándar sólidos o líquidos. Del mismo modo, la mayor parte de los equipos se pueden calibrar de manera que se conozcan de antemano los errores sistemáticos. La importancia de esta área de la química (y de otras ciencias experimentales) se refleja en el amplio trabajo de organismos como el *National Physical Laboratory* (en Inglaterra), *National Bureau of Standards* (en Estados Unidos), y organizaciones similares en otras partes del mundo. Se han escrito volúmenes completos sobre libros de la estandarización de tipos particulares de equipos (véase la bibliografía), y una serie de organizaciones comerciales se especializan en la venta de materiales de referencia estándar.

Si los errores sistemáticos se producen en el transcurso de procesos químicos o como resultado del uso de reactivos impuros, en lugar de originarse en el equipo, puede emplearse una forma de comparación alternativa, es decir, la determinación debe repetirse mediante un procedimiento completamente independiente. Si se utilizan dos (o más) métodos físicos o químicos no relacionados al efectuar un análisis, y si conducen en forma consistente a resultados que muestran sólo diferencias aleatorias, entonces es razonable suponer que no hay errores sistemáticos significativos. Resulta evidente que, para que esta aproximación sea válida, *cada paso* de los dos experimentos tiene que ser independiente. Como en el caso de la determinación de cromo en suero, no será suficiente sustituir el paso de la espectrometría de absorción atómica por un método colorimétrico o por espectrometría de plasma, es decir, los errores sistemáticos sólo se revelarían si se alteran además los métodos de muestreo, por ejemplo, eliminando o minimizando el uso del material de acero inoxidable. Otro aspecto importante es que las comparaciones deben hacerse con base en el intervalo de concentraciones total para el cual se usa un procedimiento analítico. Por ejemplo, el método del verde de bromocresol que se utiliza con frecuencia para la determinación de albúmina en suero, se correlaciona con métodos alternativos (p. ej., inmunológicos) a niveles normales o altos de albúmina; pero cuando los niveles de albúmina son anormalmente bajos (éstos son inevitablemente los casos de mayor interés clínico), la concordancia entre los dos métodos es escasa, y el método del colorante proporciona en forma consistente y errónea concentraciones de albúmina más altas. Las aproximaciones estadísticas utilizadas en las comparaciones de los métodos se describen con detalle en los capítulos 3 y 5.

El predominio de los errores sistemáticos en el trabajo analítico cotidiano se puede aclarar por medio de los resultados de los **ensayos de colaboración**. Para un analista experimentado y capaz es natural creer que sí, por ejemplo, encuentra 10 ng/ml de un fármaco en una muestra de orina, entonces otros analistas obtendrían resultados similares para la misma muestra, y las diferencias sólo se deberían a errores aleatorios. Por desgracia,



en la práctica esto se encuentra lejos de la verdad. En muchos trabajos de colaboración en los que intervienen diferentes laboratorios, incluso cuando se examinan, mediante los mismos procedimientos experimentales y los mismos tipos de experimentos, partes alícuotas de una muestra preparadas y distribuidas meticulosamente, la variación en los resultados a menudo excede en mucho lo que razonablemente se podría esperar de los errores aleatorios. La conclusión ineludible es que en muchos laboratorios no se detectan o corrigen una serie de errores sistemáticos, tanto positivos como negativos. La importancia evidente de esta situación preocupante, que tiene serias implicaciones para todos los científicos analíticos, ha alentado muchos estudios sobre la metodología de los ensayos de colaboración, y la evaluación estadística de sus resultados. Este tema se analiza con mayor detalle en el capítulo 4.

## 1.6 PLANIFICACIÓN Y DISEÑO DE EXPERIMENTOS

Muchos químicos consideran las pruebas estadísticas como métodos que sólo se utilizan para juzgar los resultados de un experimento concluido. Aunque esto de hecho es un área de aplicación esencial de la estadística, hemos visto que no vale la pena considerar resultado cuantitativo alguno si no va acompañado por una estimación de sus errores. También es necesario estar consciente de la importancia de los conceptos estadísticos en la planificación y el diseño de experimentos. En la sección anterior se subrayaba la importancia de intentar predecir errores sistemáticos, lo cual permite al analista hacer planes para después valorarlos. Las mismas consideraciones se aplican a los errores aleatorios. Como veremos en el capítulo 2, la combinación de errores aleatorios de las partes individuales de un experimento para dar un error global requiere el uso de fórmulas estadísticas sencillas. En la práctica, el error global está dominado a menudo por el error en una etapa del experimento, en tanto que otros errores muestran efectos insignificantes cuando todos los errores se combinan en forma correcta. De nueva cuenta, es obvio que se espera identificar, *antes de que empiece el experimento*, en dónde puede surgir el error dominante, y entonces intentar minimizarlo. Aunque los errores aleatorios nunca pueden eliminarse, en realidad pueden minimizarse atendiendo particularmente las técnicas experimentales: un ejemplo sencillo es el de mejorar la precisión de un experimento espectrométrico usando una célula de muestra a temperatura constante. En consecuencia, para los errores aleatorios y sistemáticos la moraleja es clara: antes de comenzar el experimento, hay que esforzarse por identificar las fuentes graves de error, de manera que el experimento pueda diseñarse para minimizar tales errores.

También debemos hacer algunas consideraciones preliminares sobre otros aspectos aún más sutiles del diseño experimental. En muchos análisis, encontramos que una o más características deseables del método (p. ej., sensibilidad, selectividad, velocidad de muestreo, bajo costo, etc.) dependen de una serie de factores experimentales. Al diseñar el experimento, hay que intentar que se use de manera adecuada la combinación óptima de estos factores, y obtener la mejor sensibilidad, selectividad, etc. Sin embargo, pueden ser necesarios *algunos* experimentos preliminares o conocimientos anteriores; la optimización deberá realizarse (con el fin de conservar recursos y tiempo) antes de que el método se utilice en forma rutinaria o se haga un uso generalizado del mismo.

La complejidad de los procedimientos de optimización puede aclararse con la ayuda de un ejemplo. En el análisis enzimático, la concentración del analito se determina a partir

de observaciones de la velocidad de una reacción catalizada por enzimas. El analito es a menudo el sustrato, es decir, el compuesto que cambia en la reacción catalizada por la enzima. Supongamos que se busca la máxima velocidad de reacción en un experimento concreto, y que la velocidad en la práctica depende (entre otros factores) del pH de la mezcla de reacción, la temperatura y la concentración de enzima. ¿Cómo se pueden encontrar las condiciones óptimas? Es fácil identificar una aproximación posible. El analista podría realizar una serie de experimentos en los cuales la concentración de enzima y la temperatura permanezcan constantes, pero se varíe el pH. En cada caso podría determinarse la velocidad de la reacción catalizada por la enzima y se podría obtener el valor del pH óptimo, supongamos que es de 7.5. Podría realizarse una segunda serie de experimentos sobre la velocidad de reacción, manteniendo constante el pH a 7.5 y fijando la concentración de enzima nuevamente, pero se variaría la temperatura; así se obtendría una temperatura óptima, supongamos de 40°C. Por último, una serie de experimentos con pH 7.5 y 40°C, pero con variaciones de enzima, indicarían el nivel óptimo de enzima. Resulta tedioso aproximarse así a la optimización del experimento: en ejemplos de investigación más reales pueden necesitarse más de tres factores experimentales. A ello se añade además otra objeción fundamental que supone que las variables o factores (pH, temperatura, concentraciones de enzima) afectan la velocidad de reacción de manera *independiente*. Esto puede no ser verdad. Por ejemplo, se ha encontrado que con un pH de 7.5 la temperatura óptima es de 40°C. Sin embargo con un pH diferente, la temperatura óptima podría no ser de 40°C. En otras palabras, estos factores pueden afectar la velocidad de reacción en una forma *interactiva*, y las condiciones establecidas en las series de experimentos que se acaban de describir pueden no ser en realidad las óptimas. Una vez comenzada la primera serie de experimentos en diferentes condiciones, se puede obtener entonces una serie diferente de valores "óptimos". Con este ejemplo sencillo queda claro que la optimización experimental puede ser un problema en extremo complejo. Este importante aspecto de la estadística, aplicado a la química analítica, se considera con más detalle en el capítulo 7.

## 1.7 CALCULADORAS Y COMPUTADORES EN LOS CÁLCULOS ESTADÍSTICOS

Ningún químico puede desconocer los sorprendentes avances que recientemente han surgido en el ámbito de la microelectrónica. Estos avances han hecho posible la construcción de aparatos que simplifican en gran medida los cálculos estadísticos. El rápido crecimiento de la "quimiometría" (la aplicación de métodos matemáticos a la solución de problemas químicos de todos los tipos) se debe a la facilidad con que pueden manejarse grandes cantidades de datos, y realizar cálculos complejos, con calculadoras y computadores.

Estos aparatos están a disposición del químico analítico en varios niveles de complejidad y costo, como se indica a continuación.

- (1) Las calculadoras de bolsillo son muy baratas, muy confiables y capaces de realizar muchos cálculos estadísticos de rutina pulsando un número mínimo de teclas. Muchas calculadoras, provistas de funciones preprogramadas, permitirán los cálculos de la media y la desviación estándar (véase el cap. 2) y de regresión lineal y correlación (véase el cap. 5). A otras calculadoras se les puede adaptar tarjetas adicionales o

programarlas el usuario para realizar otros cálculos, tales como límites de confianza (véase el cap. 2); pruebas de significación (véase el cap. 3), y regresiones no lineales (véase el cap. 5). Para muchos cálculos en laboratorios donde se realiza investigación analítica o análisis de rutina, las calculadoras de este tipo pueden resultar más adecuadas. Sus desventajas residen en la imposibilidad de manejar cantidades grandes de datos, y los problemas que surgen del limitado número de cifras significativas incorporadas a sus algoritmos (véase el cap. 2).

(2) Los microcomputadores, que en algunos casos tienen un costo ligeramente mayor al de las calculadoras, se encuentran en la actualidad en muchos laboratorios analíticos, deben programarse para realizar los cálculos estadísticos más sencillos (aunque se encuentran disponibles en el mercado paquetes de software estadístico), pero pueden manejar grandes cantidades de datos. Por lo regular utilizan más cifras que las calculadoras, aunque a veces también se presentan problemas de este tipo durante los cálculos estadísticos. Los microcomputadores pueden convertirse también, con la ayuda de un programa de computador, en excelentes procesadores de textos (este libro se escribió y se tradujo mediante este sistema!), y esta posibilidad es de mucha ayuda para elaborar informes sobre trabajos de laboratorio y documentos similares. Otros programas muy valiosos son las "hojas de cálculo", las cuales fueron diseñadas inicialmente para el mundo de los negocios y las finanzas, pero las posibilidades de realizar cálculos muy complejos de hileras y columnas de números, pueden adaptarse con facilidad a cálculos estadísticos tales como las regresiones lineales (véase el Cap. 5). Quizás la mayor importancia de los microcomputadores es que pueden interconectarse con facilidad a casi todo tipo de equipos electrónicos de laboratorio. Entonces el microcomputador puede utilizarse para optimizar y controlar las operaciones del instrumento, recoger, grabar y procesar todos los datos resultantes, y producir una copia impresa de los resultados calculados. Pueden conectarse varios instrumentos a un único computador y, mediante el uso de sistemas de redes, pueden conectarse varios computadores a un almacén de datos centralizado, como un sistema de disco duro. Los microcomputadores pueden convertirse también en terminales de un computador central cuando se necesita manejar una gran cantidad de cálculos complejos.

(3) Muchos instrumentos de laboratorio llevan incorporados sus propios microcomputadores. Estos sistemas se diseñan específicamente para controlar el instrumento de que se trata por medio de funciones preprogramadas. No poseen la flexibilidad de los sistemas del microcomputador (a menudo el usuario no se da cuenta de las técnicas de programación y de los algoritmos de cálculo utilizados, y no es capaz de cambiarlos), pero pueden realizar cálculos de alta complejidad y operaciones de control. Un avance reciente en esta área es el desarrollo de los denominados instrumentos "inteligentes" en los que el microprocesador de control (o microcomputador) es capaz de calibrar el instrumento, de diagnosticar funciones incorrectas y corregirlas, y de optimizar las especificaciones de un instrumento. Dichos desarrollos continúan mejorando, y aumentan en forma progresiva la eficiencia de los laboratorios analíticos a un costo relativamente bajo. Las investigaciones recientes en sistemas expertos y en los métodos de inteligencia artificial producirán instrumentos que en realidad proporcionen a los usuarios "consejos" sobre los mejores parámetros de operación, o fijen estos parámetros de manera automática a la luz de la introducción de datos muestrales y de la "experiencia" del instrumento. De nueva cuenta hay muchas razones para pensar

que dichos sistemas sofisticados se pueden basar en microprocesadores o microcomputadores.

(4) Grandes computadores, por ejemplo los denominados minicomputadores y computadores centrales, se utilizan ahora únicamente para el control de instrumentos muy grandes y complejos, o para cálculos especializados, tales como los que usan bases de datos extensas o que requieren muchas iteraciones. Sin embargo, parece que en cuanto se mejore la velocidad y capacidad de memoria de los computadores personales, disminuirá el uso de grandes aparatos en los laboratorios analíticos.

Es muy importante para el químico analítico recordar que la disponibilidad de todos estos instrumentos de manejo de datos aumenta, en vez de disminuir, la necesidad de un sólido conocimiento de los principios en los que se basan los cálculos estadísticos. Un computador o calculadora realizará con rapidez cualquier prueba estadística o cálculo seleccionado por el usuario, *tanto si ese procedimiento es adecuado o no para los datos en estudio*. Por ejemplo, un programa de mínimos cuadrados lineal determinará una línea recta sea cual sea la serie de valores de  $x$  y  $y$  que se pretenda ajustar, aún en los casos en que la inspección visual muestre de inmediato que dicho programa es inadecuado (véase el Cap. 5). De nueva cuenta, un programa sencillo para probar la significación de la diferencia entre las medidas de dos series de datos puede suponer que las varianzas (véase el Cap. 2) de las dos series sean similares; no obstante, el programa realizará a ciegas los cálculos requeridos y proporcionará un "resultado" aun cuando las varianzas difieran significativamente. Incluso, series muy completas de programas de computador suelen ser incapaces de proponer método estadístico apropiado para un conjunto de datos específicos. Entonces, el analista debe usar sus conocimientos de estadística, y su sentido común, para asegurar que se realiza el cálculo adecuado. Como ya hemos señalado, cuando los microprocesadores están incorporados en los instrumentos analíticos, el usuario a menudo ignora la naturaleza exacta de los cálculos realizados con los datos experimentales. Ésta es una situación indeseable, aunque la renuencia de los fabricantes de instrumentos a divulgar programas de computador de difícil desarrollo es bastante comprensible.

En este libro no se pretende de ninguna manera proporcionar listas completas de programas de computador para cálculos estadísticos. En el mercado se encuentran disponibles programas excelentes para muchos de los métodos que se describen en los capítulos siguientes, incluyendo las técnicas más avanzadas como la regresión no lineal (Cap. 5); el reconocimiento de pautas, y la optimización simplex (Cap. 7). En general, se debe elegir cuidadosamente paquetes de los programas para adecuarlos a la marca y el modelo del computador que se vaya a utilizar.

## BIBLIOGRAFÍA

1. Muchos libros de texto convencionales de química analítica incluyen secciones o capítulos referentes al tratamiento estadístico de resultados experimentales. Estas secciones son muy variadas en su cobertura, aunque un buen ejemplo es el capítulo 3 de *Fundamentals of Analytical Chemistry* de Skoog D. A. y West D. M., 4a ed., Nueva York, Holt Saunders, 1982.
2. Se han escrito muchos libros de estadística general. Entre ellos, *Facts from Figures* de Moroney M. J., Penguin Books, es un documento clásico atractivo y de mucho prestigio. Una publicación más reciente y que merece ser leída, es *Statistics in Action* de Sprent, Peter, Penguin Books. Un

buen número de textos estadísticos están dirigidos a usuarios científicos (no necesariamente químicos); *Statistics for Technology* de Chatfield, Christopher, Chapman & Hall, Londres, es un ejemplo excelente. Tales libros proporcionan un contenido teórico general, muy asequible, de muchos de los métodos que se describen en este volumen.

3. Todos los usuarios de la estadística requieren una buena serie de tablas estadísticas. Se recomienda *Elementary Statistics Tables* de Neave, Henry R., Allen & Urbis, Londres, por su clara exposición de tablas y de las pruebas estadísticas asociadas.
4. Varios cuerpos estatutarios y profesionales en Inglaterra, Estados Unidos y otros países publican manuales y monografías sobre procedimientos estadísticos recomendados y estándares, y sobre materiales estándar. Como ejemplos se incluyen al British Standards Institute (p. ej., BS 5497, *Precision of Tests Methods, Part 1 - Guide to the Determination of Repeatability and Reproducibility of a Standard Test Method* y BS 1797, *Calibration of Volumetric Glassware*), la Association of Official Analytical Chemists (Youden, W. J. y Steiner, E. H., *Statistical Manual of the AOAC*, 1975), la International Organisation for Standardisation (ISO) (p. ej., IS4259, *Petroleum Products - Determination and Application of Precision Data in Relation to Methods of Tests*) y la American Society for Testing and Materials (ASTM) (p. ej. Standard Specifications E617-78, *Laboratory Weights and Precision Mass Standards*, y E694-79, *Volumetric Ware*).
5. Las monografías que describen procedimientos estándar y materiales en áreas específicas de estudio están ejemplificadas por Knowles, A. y Burgess, C. Eds., *Standards in Absorption Spectrometry*, y Miller, J. N. Ed., *Standards in Fluorescence Spectrometry*, ambas publicadas por Chapman & Hall en conjunto con el grupo de espectrometría de UV. La National Bureau of Standards de Estados Unidos publica monografías que acompañan a su extenso catálogo de materiales estándar. Un ejemplo reciente, que describe el uso de sulfato de quinina como estándar en espectrometría de fluorescencia, es la Publicación Especial de la NBS No 260-64 1980 de Velapoli, R. A. y Mielenz, K. D.
6. Recientemente se encuentran disponibles varios libros que describen métodos estadísticos y proporcionan listas de programas de computador en BASIC. Estos incluyen dos libros de Lee, J. D. y Lee, T. D., *Statistics and Computer Methods in BASIC* y *Statistics and Numerical Methods in BASIC for Biologists*, ambos publicados por Van Nostrand Reinhold; *Basic Statistical Computing* de Cooke, D., Craven, A. H. y Clarke, G. M., publicado por Arnold; *BASIC Microcomputing and Biostatistics* de Rogers, D. W., publicado por Humana Press Inc.

## EJERCICIOS

1. Una muestra estándar de suero sanguíneo humano contiene 42.0 g de albúmina por litro. Cinco laboratorios (A-E) realizan cada uno seis determinaciones (en el mismo día) de la concentración de albúmina, con los siguientes resultados (en g/l):

A	42.5	41.6	42.1	41.9	41.1	42.2
B	39.8	43.6	42.1	40.1	43.9	41.9
C	43.5	42.8	43.8	43.1	42.7	43.3
D	35.0	43.0	37.1	40.5	36.8	42.2
E	42.2	41.6	42.0	41.8	42.6	39.0

Comente la exactitud y precisión de cada una de estas series de resultados.

2. Utilizando la misma muestra y método del ejercicio 1, el laboratorio A realiza otras seis determinaciones de la concentración de albúmina en un tiempo de seis días sucesivos. Los valores obtenidos son 41.5, 40.8, 43.3, 41.9, 42.2, 41.7 g/l. Comente estos resultados.

3. Se ha determinado cuatro veces el número de lugares de unión por molécula en una muestra de anticuerpos monoclonados, con resultados de 1.95, 1.95, 1.92, 1.97. Comente la exactitud y la precisión de estos resultados.

4. Analice el grado de exactitud y precisión deseables en los siguientes análisis:

- (i) Determinación de la concentración de lactato en muestras de sangre humana.
- (ii) Determinación de uranio en una muestra de mineral.
- (iii) Determinación de una droga en plasma sanguíneo después de una sobredosis.
- (iv) Estudio de la estabilidad de un reactivo colorimétrico por determinación de su absorbancia a una longitud de onda durante un periodo de varias semanas.

5. Para cada uno de los siguientes experimentos, intente identificar las principales fuentes probables de error sistemático y aleatorio, y plantee de qué manera pueden minimizarse dichos errores:

- (i) El contenido de hierro de un gran trozo de mineral se determina tomando una pequeña muestra, disolviéndola en ácido, y titulando con sulfato cérico después de reducir el Fe(III) a Fe(II).
- (ii) Se utiliza el mismo muestreo y procedimiento de disolución de (i), aunque el hierro se determina colorimétricamente después de añadir un reactivo quelante y de extraer al complejo coloreado resultante en un disolvente orgánico.
- (iii) El contenido de sulfato en una solución acuosa se determina gravimétricamente con cloruro bórico como precipitante.

## Errores en el análisis clásico. Estadística de medidas repetidas

### 2.1 MEDIA Y DESVIACIÓN ESTÁNDAR

En este capítulo se introducen algunos conceptos estadísticos fundamentales y se aplican a situaciones habituales en el análisis clásico es decir, se realizan mediciones repetidas de la misma cantidad. En el capítulo 1 se expusieron los diversos tipos de errores al considerar los resultados de repetir cinco veces una valoración hecha por cuatro estudiantes: estos datos se reproducen a continuación.

Estudiante	Resultados (ml)				
A	10.08	10.11	10.09	10.10	10.12
B	9.88	10.14	10.02	9.80	10.21
C	10.19	9.79	9.69	10.05	9.78
D	10.04	9.98	10.02	9.97	10.04

Para comparar estos resultados se utilizaron dos criterios, el valor promedio y el grado de variabilidad (dispersión). El valor promedio utilizado fue la **media aritmética**  $\bar{x}$ , (en forma abreviada, **media**) que es la suma de todas las mediciones dividida por el número de mediciones:

$$\bar{x} = \sum_i x_i / n \quad (2.1)$$

La medida más utilizada de la variabilidad es la **desviación estándar**,  $s$ . Ésta se define por la fórmula:

$$s = \sqrt{\sum_i (x_i - \bar{x})^2 / (n - 1)} \quad (2.2)$$

El cálculo de la desviación estándar de los resultados de A se muestra en la tabla 2.1. El lector debería comprobar que las desviaciones estándar de los resultados de los estudiantes

B, C y D son 0.17, 0.21 y 0.033 ml respectivamente, y establecer así una confirmación cuantitativa de las afirmaciones sobre la precisión realizadas en el capítulo 1.

Muchas calculadoras de bolsillo proporcionan los resultados de estos cálculos si se introducen en ellas los valores de  $\bar{x}$ . Sin embargo, se debe tener cuidado en pulsar la tecla correcta para obtener la desviación estándar. Algunas calculadoras proporcionan dos valores diferentes para la desviación estándar, uno calculado utilizando la ecuación (2.2) y el otro sustituyendo  $(n - 1)$  por  $n$  en el denominador de esta ecuación. (La razón de estas dos formas diferentes se explica en la pág. 23.) Por supuesto, para valores grandes de  $n$  la diferencia es despreciable.

Por desgracia, al calcular una desviación estándar, la calculadora puede redondear números de manera tal que puede obtenerse un valor erróneo (incluso cero). Esto sucede normalmente cuando existen diferencias entre los valores introducidos en el cuarto o subsiguientes dígitos significativos, dependiendo del tipo de calculadora. Por ejemplo, muchas dan como cero la desviación estándar de los tres valores 100.000, 100.001 y 100.002, mientras que de hecho es 0.001. Este problema se puede superar **codificando** los valores, es decir, restando una cantidad fija a cada uno, por ejemplo, en este caso 100.000 para obtener 0.000, 0.001 y 0.002. Puesto que la desviación estándar mide la variabilidad en torno a la media, la desviación estándar de esos valores codificados es la misma que la de los valores originales. (La media de los valores originales se obtiene añadiendo 100 a la media de los valores codificados.) La codificación de valores también reduce el trabajo y posibles errores al introducir la información en una calculadora. En este ejemplo concreto, el cálculo podría haberse realizado aún con más rapidez y con menos tendencia al error tomando los valores codificados como 0, 1 y 2: en tal caso debe tenerse especial cuidado de decodificar el resultado que aparece en la pantalla de la calculadora.

**Tabla 2.1** Cálculos de la media y la desviación estándar para los resultados del estudiante A

	$x_i$ (ml)	$(x_i - \bar{x})^2$
	10.08	0.0004
	10.11	0.0001
	10.09	0.0001
	10.10	0.0000
	10.12	0.0004
Total	50.50	0.0010
		$\bar{x} = 50.50/5 = 10.10 \text{ ml}$
		$s = \sqrt{0.0010/4} = 0.016 \text{ ml}$

El cálculo de la tabla 2.1 fue muy simple porque los valores de  $(x_i - \bar{x})^2$  se podrían calcular mentalmente. Por lo regular esto no es así y para simplificar la aritmética se puede utilizar una forma alternativa de la ecuación (2.2), si no se dispone de una calculadora programada para cálculos estadísticos:

$$s = \sqrt{\frac{\sum_i x_i^2}{(n - 1)} - \frac{(\sum_i x_i)^2}{n(n - 1)}} \quad (2.3)$$

La media y la desviación estándar pueden calcularse también con un computador utilizando un programa BASIC de unas pocas líneas. Este programa sólo sería útil si el propio computador fuese recogiendo los datos o si se analizase una gran cantidad de ellos. Ejemplos de dichos programas aparecen en los libros de la bibliografía del capítulo 1.

El cuadrado de  $s$  es una cantidad estadística muy importante conocida como la **varianza**; su significación se manifestará en el capítulo dedicado a la propagación de errores. Así mismo se emplea mucho el **coeficiente de variación (CV)**, también conocido como la **desviación estándar relativa (DER)**, que se define por  $100 s/\bar{x}$ . El CV o DER, cuyas unidades se expresan obviamente en porcentaje, es un ejemplo de **error relativo**; es decir, una estimación del error dividida por una estimación del valor absoluto de la cantidad medida. Los errores relativos se utilizan con frecuencia en la comparación de las precisiones de los resultados que tienen diferentes unidades o magnitudes, y son importantes en el cálculo de la propagación de errores.

2.2 DISTRIBUCIÓN DE ERRORES

Aunque la desviación estándar proporciona una medida de la dispersión de un conjunto de resultados alrededor del valor medio, no indica la forma en que están distribuidos los resultados. Para aclarar esto se requiere un gran número de mediciones como las expuestas en la tabla 2.2, en la cual se proporcionan los resultados de 50 determinaciones repetidas

Tabla 2.2 Resultados de 50 determinaciones de concentración de ion nitrato, en  $\mu\text{g/ml}$

0.51	0.51	0.51	0.50	0.51	0.49	0.52	0.53	0.50	0.47
0.51	0.52	0.53	0.48	0.49	0.50	0.52	0.49	0.49	0.50
0.49	0.48	0.46	0.49	0.49	0.48	0.49	0.49	0.51	0.47
0.51	0.51	0.51	0.48	0.50	0.47	0.50	0.51	0.49	0.48
0.51	0.50	0.50	0.53	0.52	0.52	0.50	0.50	0.51	0.51

de la concentración de ion nitrato, dada con dos cifras significativas, en una muestra concreta de agua. Estos resultados se pueden resumir en una **tabla de frecuencias** (tabla 2.3).

Tabla 2.3 Tabla de frecuencias para las mediciones de concentración de ion nitrato

Concentración de ion nitrato ( $\mu\text{g/ml}$ )	Frecuencia
0.46	1
0.47	3
0.48	5
0.49	10
0.50	10
0.51	13
0.52	5
0.53	3

Esta tabla indica que, en la tabla 2.2, el valor de  $0.46 \mu\text{g/ml}$  aparece una vez, el valor de  $0.47 \mu\text{g/ml}$  aparece tres veces y así sucesivamente. El lector puede comprobar que la media de estos resultados es  $0.500 \mu\text{g/ml}$  y la desviación estándar es  $0.0165 \mu\text{g/ml}$ . (Estos valores se expresan de manera arbitraria con tres cifras significativas: este importante aspecto de la presentación de resultados se analiza al final del capítulo.) La distribución de los resultados puede apreciarse con facilidad dibujando un **histograma** como el de la figura 2.1. En él se muestra que las mediciones están distribuidas en forma casi simétrica en torno a la media, con las mediciones agrupadas respecto al centro.

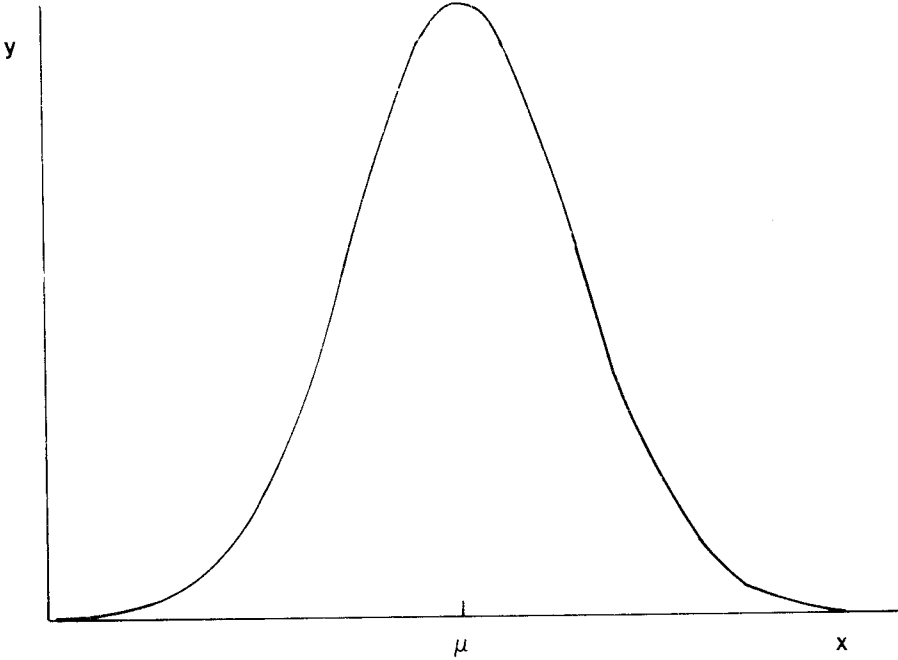


Figura 2.1 Histograma de los datos de concentración de ion nitrato de la tabla 2.3.

Este conjunto de 50 mediciones constituye una **muestra** de un gran número de ellas (en teoría infinita) que se pueden tomar de la concentración de nitrato. Este conjunto de posibles mediciones se denomina **población**. Si no hay errores sistemáticos, entonces la media de la población, indicada por  $\mu$  es el verdadero valor de la concentración de nitrato que se intenta determinar. La media,  $\bar{x}$ , de la muestra nos proporciona una estimación de  $\mu$ . De manera similar, la población tienen una desviación estándar simbolizada por  $\sigma$ . El valor de la desviación estándar,  $s$ , de la muestra nos proporciona una estimación de  $\sigma$ . El uso de la ecuación (2.2) nos da una estimación insesgada de  $\sigma$ . Si en vez de  $(n - 1)$  utilizamos  $n$  en el denominador de la ecuación, el valor que se obtiene de  $s$  tiende a subestimar  $\sigma$  (véase la pág. 21).

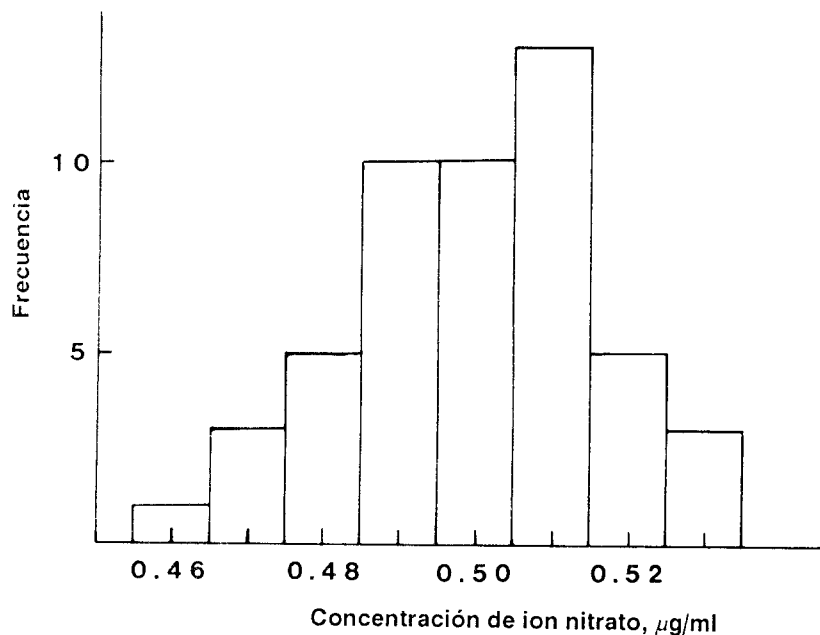
Las mediciones de concentración de ion nitrato que aparecen en la tabla 2.2 presentan sólo valores discretos debido a las limitaciones del método de medición. En teoría, una concentración puede tomar cualquier valor de manera que para describir la forma de la población, de la que se ha extraído una muestra, se necesita una curva continua. El modelo



matemático que habitualmente se emplea es la **distribución normal** o **Gaussiana** descrita por la ecuación

$$y = \frac{\exp\{-(x - \mu)^2/2\sigma^2\}}{\sigma\sqrt{2\pi}} \quad (2.4)$$

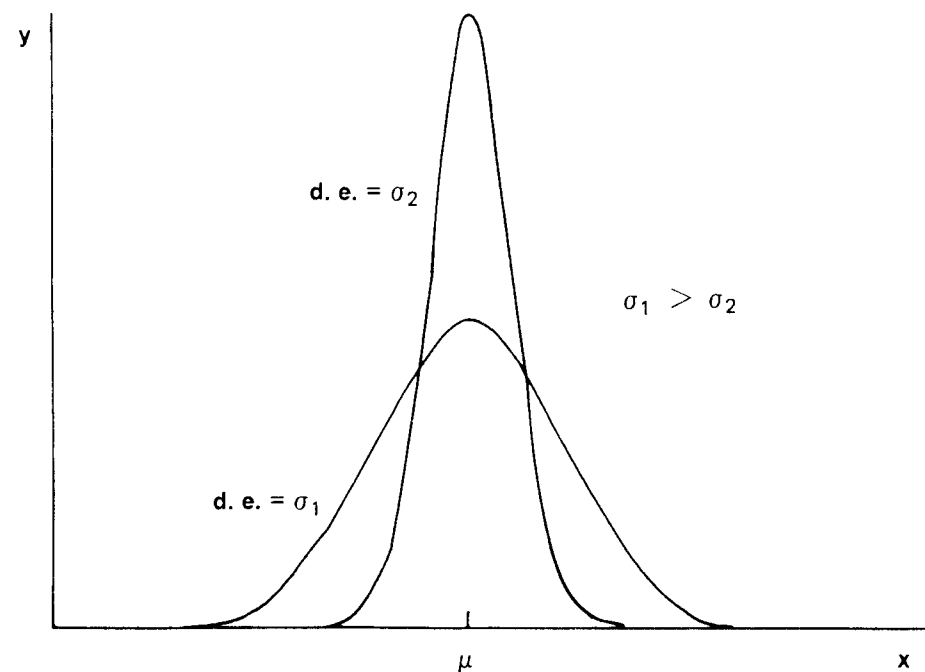
y su forma se muestra en la figura 2.2. No es necesario recordar esta fórmula complicada, aunque algunas de sus propiedades generales son importantes. La curva es simétrica respecto a  $\mu$  y cuanto mayor sea el valor de  $\sigma$  mayor será la variabilidad de la curva, como



**Figura 2.2** La distribución normal,  $y = \exp[-(x - \mu)^2/2\sigma^2]/\sigma\sqrt{2\pi}$ . La media se indica con  $\mu$ .

se muestra en la figura 2.3. Un análisis un poco más detallado demuestra que, cualesquiera que sean los valores de  $\mu$  y  $\sigma$ , aproximadamente el 68% de los valores de la población caen dentro de  $\pm 1\sigma$  de la media; cerca del 95% de los valores se ubican dentro de  $\pm 2\sigma$  de la media, y casi el 99.7% de los valores se encuentran dentro de  $\pm 3\sigma$  de la media (figura 2.4). Esto significa que si las concentraciones de ion nitrato que aparecen en la tabla 2.3 se distribuyen normalmente, cerca del 68% caerían en el intervalo de 0.483 a 0.517; alrededor del 95% en el intervalo de 0.467 a 0.533, y el 99.7% en el de 0.450 a 0.550. De hecho, 33 de los 50 resultados (66%) caen entre 0.483 y 0.517; 49 (98%) entre 0.467 y 0.533, y todos los resultados se ubican entre 0.450 y 0.550, de manera que la concordancia con la teoría es bastante satisfactoria.

Aunque no se pueda demostrar que las mediciones repetidas de cualquier cantidad analítica siempre van a estar distribuidas normalmente, las pruebas indican que por lo general esta hipótesis está al menos muy cerca de la verdad. Además, como veremos al



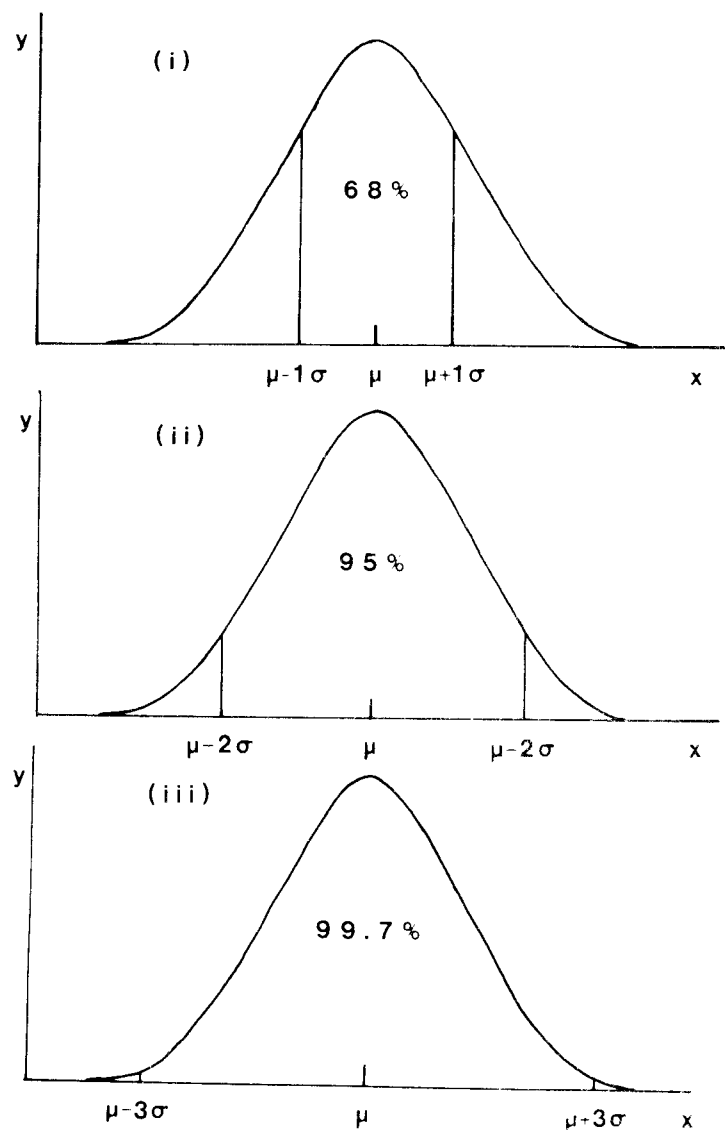
**Figura 2.3** Distribuciones normales con la misma media pero diferentes valores de desviación estándar.

estudiar las medias muestrales, cualquier desviación de la normalidad en una población no es generalmente importante en el contexto de las pruebas estadísticas utilizadas con más frecuencia.

La distribución normal no sólo se aplica cuando se hacen mediciones repetidas de un mismo espécimen; también los resultados obtenidos a menudo se adaptan a la distribución normal cuando se mide la misma magnitud para diferentes muestras. Por ejemplo, si medimos las concentraciones de albúmina en suero sanguíneo procedentes de humanos adultos sanos, encontraremos que los resultados estarán aproximadamente distribuidos de manera normal. Sin embargo, en este segundo tipo de población, es decir, una medida sobre cada serie de especímenes, no es raro que existan otras distribuciones. En concreto, se encuentra con frecuencia la denominada **distribución log-normal**: en esta distribución los *logaritmos* de las concentraciones (u otras características), cuando se representan frente a la frecuencia, proporcionan una curva de distribución normal. Por ejemplo, la concentración de anticuerpos en suero sanguíneo humano es aproximadamente una distribución log-normal (figura 2.5), y los tamaños de la partícula de gotas formadas por nebulizadores en fotometría de llama pueden seguir también esta distribución.

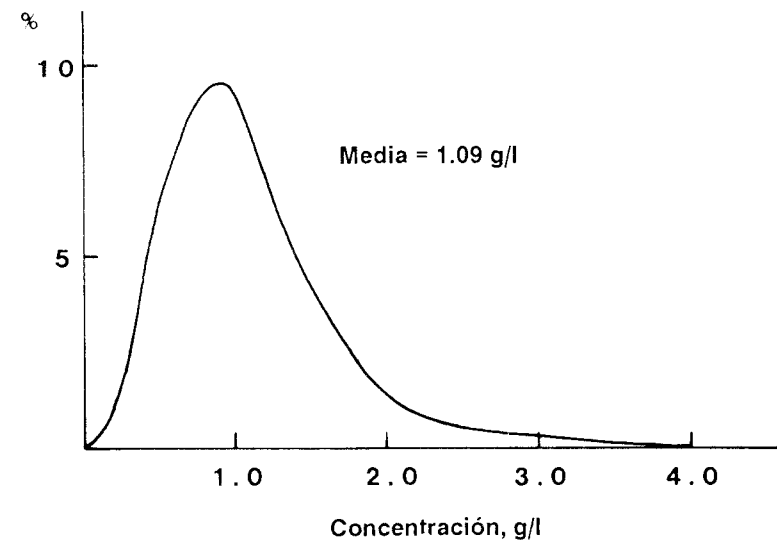
## 2.3 LA DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Ya hemos visto que la media de una muestra de mediciones nos proporciona una estimación del valor verdadero,  $\mu$ , de la magnitud que se quiere medir. Sin embargo, ya que las



**Figura 2.4** Propiedades de la distribución normal: (i) aproximadamente el 68% de los valores caen dentro de  $\pm 1\sigma$  de la media; (ii) cerca del 95% de los valores se ubican dentro de  $\pm 2\sigma$  de la media; (iii) el 97% de los valores se encuentran dentro de  $\pm 3\sigma$  de la media.

mediciones individuales están distribuidas en torno al valor verdadero con una dispersión que depende de la precisión, es poco probable que la media de la muestra sea *exactamente* igual al valor verdadero. Por esta razón es más útil proporcionar un intervalo de valores que contenga casi con seguridad el valor verdadero. La amplitud de este intervalo depende de dos factores. El primero es la precisión de las mediciones individuales, las cuales dependen a su vez de la varianza de la población. El segundo es el número de mediciones de la muestra. El sólo hecho de repetir mediciones implica que se tiene más confianza en



**Figura 2.5** Una distribución aproximadamente log-normal: concentración de anticuerpos de inmunoglobulina M en suero de varones. Observe la asimetría de la curva.

la media de varios valores que en uno solo. Mucha gente piensa que cuantas más mediciones se tomen más confiable será nuestra estimación de  $\mu$ , el valor real. Para corroborar esta idea volvamos a la determinación de ion nitrato descrita en la sección 2.2. En casos como el estudiado (concentración del ion nitrato), por lo regular es poco práctico realizar 50 mediciones repetidas. Un número más verosímil sería 5 y podemos ver cómo las mediciones de las muestras de este tamaño están distribuidas alrededor de  $\mu$ , si consideramos los resultados de la tabla 2.2 como 10 muestras en las que cada una de ellas contiene 5 resultados. Si tomamos cada columna como una muestra, las medias son 0.506, 0.504, 0.502, 0.496, 0.502, 0.492, 0.506, 0.504, 0.500, 0.486. Resulta evidente que estas mediciones se encuentran más agrupadas entre sí que las mediciones originales. De la misma forma que éstas eran muestras de una población infinita de posibles mediciones, estas medias son una muestra de las posibles medias de muestras de 5 mediciones de la población global. La distribución de estas mediciones muestrales se denomina **distribución muestral de la media o distribución en el muestreo de la media**; su media es la misma que la de la población original y su desviación estándar se denomina **error estándar de la media (e.e.m.)**. Existe una relación matemática entre el e.e.m. y la desviación estándar,  $\sigma$ , de la distribución de mediciones individuales, la cual es independiente de la forma en que están distribuidas. Si  $n$  es el tamaño de la muestra, la relación es:

$$\text{e.e.m.} = \sigma / \sqrt{n} \quad (2.4)$$

Como es de esperar, a mayor  $n$ , menor será la dispersión de las mediciones muestrales en torno a  $\mu$ . Este término utilizado universalmente, el *error estándar de la media*, podría conducir al lector a pensar que  $\sigma / \sqrt{n}$  mide la diferencia entre  $\bar{x}$  y  $\mu$ . Esto no es así:  $\sigma / \sqrt{n}$  proporciona una medida de la incertidumbre que hay al estimar  $\mu$  a partir de  $\bar{x}$ , como se verá en la siguiente sección.

Otra propiedad de la distribución muestral de la media es que, *aun cuando la población original no esté distribuida normalmente*, tiende a la distribución normal cuando aumenta  $n$ . Este resultado se conoce como **teorema del límite central**. Este teorema es de suma importancia, ya que muchas pruebas estadísticas se realizan sobre la media y se supone que se distribuye en forma normal. Debido a que en la práctica se puede suponer una distribución casi normal para las distribuciones de las mediciones repetidas, es razonable asumir que las medias de muestras muy pequeñas (digamos  $> 5$ ) se distribuyan normalmente.

## 2.4 LÍMITES DE CONFIANZA DE LA MEDIA

Ahora que conocemos la forma de la distribución muestral de la media, podemos regresar al problema de utilizar una muestra para definir el intervalo dentro del cual podamos suponer de manera razonable que se encuentra el valor verdadero. (Recuerde que al hacer esto se supone que no existen errores sistemáticos.) Este intervalo se conoce como el **intervalo de confianza** y los valores extremos de dicho intervalo se llaman **límites de confianza**. El término “confianza” implica que podemos afirmar con un grado de confianza dado, es decir, con una cierta probabilidad, que el intervalo de confianza *sí* incluye al valor verdadero. Por supuesto, el tamaño del intervalo de confianza dependerá de la certeza que queramos tener de que se incluya el valor verdadero: cuanto más grande sea la certeza, *más grande* será el intervalo requerido.

En la figura 2.6 aparece la distribución muestral de la media para muestras de tamaño  $n$ . Si suponemos a partir de ahora que la distribución es normal, entonces el 95% de las mediciones muestrales se encontrarán en el intervalo dado por:

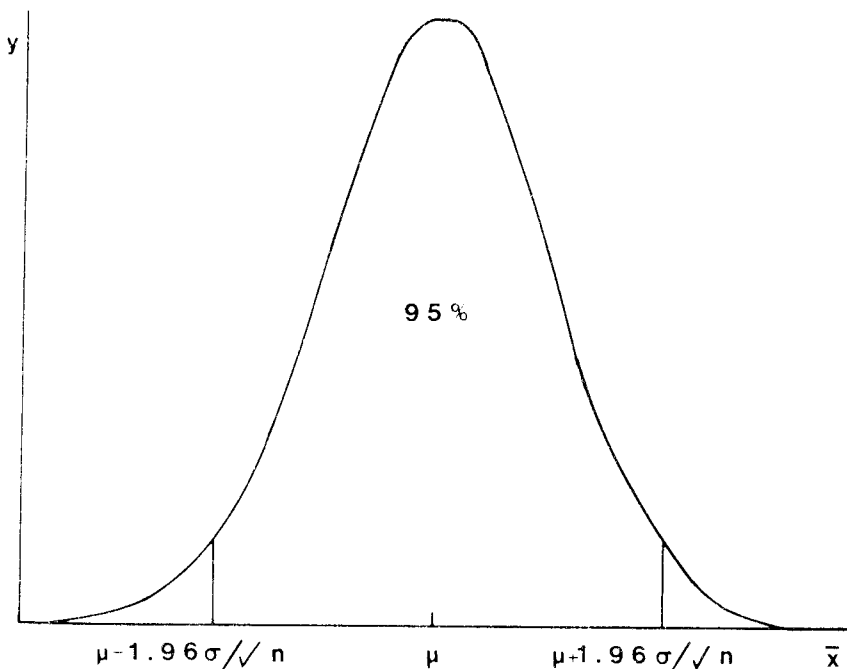


Figura 2.6 Distribución muestral de la media, en la que se muestra el intervalo dentro del cual se encuentra el 95% de las mediciones muestrales.

$$\mu - 1.96(\sigma/\sqrt{n}) < \bar{x} < \mu + 1.96(\sigma/\sqrt{n}) \quad (2.5)$$

(Se utiliza en esta ecuación el valor exacto de 1.96 en lugar del valor aproximado, 2, que se emplea frecuentemente.)

Sin embargo, en la práctica por lo regular disponemos de una muestra, de media conocida, y buscamos un intervalo para  $\mu$ , el valor verdadero. Puede reordenarse la ecuación (2.5) y expresarse de la siguiente forma:

$$\bar{x} - 1.96(\sigma/\sqrt{n}) < \mu < \bar{x} + 1.96(\sigma/\sqrt{n}) \quad (2.6)$$

La ecuación (2.6) proporciona el **intervalo de confianza al 95%** de la media. De manera similar, el intervalo de confianza al 99.7% está dado por:

$$\bar{x} - 2.97(\sigma/\sqrt{n}) < \mu < \bar{x} + 2.97(\sigma/\sqrt{n}) \quad (2.7)$$

Otro intervalo de confianza que se utiliza con frecuencia es el intervalo de confianza al 99% que está dado por:

$$\bar{x} - 2.58(\sigma/\sqrt{n}) < \mu < \bar{x} + 2.58(\sigma/\sqrt{n}) \quad (2.8)$$

La ecuación (2.6) se puede utilizar para calcular los límites de confianza al 95% para la concentración de ion nitrato. Se tiene que  $\bar{x} = 0.500$  y  $n = 50$ . Lo único que se desconoce en la ecuación es  $\sigma$ . Para muestras *grandes* como ésta,  $s$  proporciona una estimación suficientemente exacta de  $\sigma$  y puede sustituirse por ésta. De este modo, el intervalo de confianza al 95% para la concentración de ion nitrato es:

$$0.500 - 1.96 \times 0.0165/\sqrt{50} < \mu < 0.500 + 1.96 \times 0.0165/\sqrt{50}$$

y los límites de confianza resultantes son:

$$\mu = 0.5000 \pm 0.0046 \text{ } \mu\text{g/ml}$$

A medida que el tamaño de la muestra se hace más pequeño, la incertidumbre al utilizar  $s$  para estimar  $\sigma$  aumenta. A fin de permitir esto, la ecuación utilizada para calcular los límites de confianza se modifica por la siguiente:

$$\mu = \bar{x} \pm t(s/\sqrt{n}) \quad (2.9)$$

El valor adecuado de  $t$  depende tanto de  $(n - 1)$ , que se conoce como número de **grados de libertad** (cuyo símbolo habitual es  $\nu$ ), como del grado de confianza requerido. [El término “grados de libertad” se refiere al número de desviaciones *independientes* ( $x_i - \bar{x}$ ) que se utilizan al calcular  $s$ . En este caso, dicho número es  $(n - 1)$ , porque cuando se conocen  $(n - 1)$  desviaciones, la última se puede deducir si se utiliza el resultado obvio  $\sum_i (x_i - \bar{x}) = 0$ ]. En la tabla 2.4 se recogen los valores de  $t$ . En dicha tabla puede apreciarse que para muestras de tamaño superior a 50, los valores de  $t$  son muy próximos a los valores de 1.96 y de 2.58 utilizados en las ecuaciones (2.6) y (2.8), respectivamente. Esto confirma la validez de la

hipótesis utilizada antes al calcular los límites de confianza para la concentración de nitrato. El uso de esta tabla puede aclararse con un ejemplo.

*Ejemplo.* Se determinó el contenido de ion sodio de una muestra de orina utilizando un electrodo selectivo de iones, y se obtuvieron los siguientes valores: 102, 97, 99, 98, 101, 106 mM. ¿Cuáles son los límites de confianza al 95 y 99%, para la concentración de ion sodio?

La media y la desviación estándar de estos valores son 100.5 mM y 3.27 mM, respectivamente. Hay seis mediciones y por lo tanto 5 grados de libertad. El valor de *t* calculado usando la tabla 2.4 para determinar los límites de confianza al 95% es de

Tabla 2.4 Valores de *t* para intervalos de confianza

Grados de libertad	Valores de <i>t</i> para intervalos de confianza de	
	95%	99%
1	12.71	63.66
2	4.30	9.92
3	3.18	5.84
4	2.78	4.60
5	2.57	4.03
10	2.23	3.17
20	2.09	2.85
30	2.04	2.75
50	2.01	2.68
100	1.98	2.63

2.57 y, de la ecuación (2.9), los límites de confianza al 95% están dados por:

$$\begin{aligned}\mu &= 100.5 \pm 2.57 \times 3.27/\sqrt{6} \\ &= 100.5 \pm 3.4 \text{ mM}\end{aligned}$$

De igual manera, los límites de confianza al 99% están dados por:

$$\begin{aligned}\mu &= 100.5 \pm 4.03 \times 3.27/\sqrt{6} \\ &= 100.5 \pm 5.4 \text{ mM}\end{aligned}$$

Si la muestra procede de una población de la que se sabe que se distribuye log-normal, entonces los valores individuales de deberían transformar tomando sus logaritmos antes de calcular los límites de confianza. Puesto que los valores transformados estarán distribuidos normalmente, entonces los límites de confianza de la media se pueden calcular como se ha explicado con anterioridad. Sin embargo, observe que los valores de la media y la desviación típica de la muestra se calculan utilizando los valores originales no transformados. Como resultado de la asimetría de la distribución log-normal (véase la Fig. 2.5), el intervalo de confianza de la media, calculado como se ha descrito anteriormente, no es simétrico con respecto a la media muestral.

## 2.5 PRESENTACIÓN DE RESULTADOS

Como ya lo señalamos, los resultados experimentales cuantitativos carecen de interés si no van acompañados de una estimación de los errores ocurridos en su medida. Una práctica usual en la bibliografía de química analítica es citar la media como la estimación de la cantidad medida y la desviación estándar como la estimación de la precisión. Menos frecuente es citar el error estándar de la muestra en lugar de la desviación estándar, es decir, dando el resultado en la forma de los límites de confianza al 95% de la media. Ya que no existe un convenio universal, es importante establecer la forma empleada y, *suponiendo que el valor de *n* está dado*, las tres formas se pueden convertir unas en otras utilizando las ecuaciones (2.4) y (2.9).

Un aspecto que se relaciona con la presentación de resultados es el redondeo de la respuesta. El principio importante en este caso es que *el número de cifras significativas dadas indica la precisión del experimento*. Por ejemplo, sería absurdo dar el resultado de un análisis volumétrico como 0.107 846*M*, pues ningún analista podría alcanzar la precisión implicada de 0.000 001 en una cantidad medida de 0.1, es decir, 0.001%. En la práctica, se acostumbra fijar como **cifras significativas** todos los dígitos que sean seguros, más el primero incierto. Por ejemplo, la media de los valores 10.09, 10.11, 10.09, 10.10 y 10.12, es 10.102, y su desviación estándar es 0.013 04. Por supuesto, existe incertidumbre en la segunda cifra decimal; los resultados son todos 10.1 con una cifra decimal, pero difieren en la segunda. Mediante el método sugerido, el resultado podría expresarse como:

$$\bar{x} \pm s = 10.10 \pm 0.01 \quad (n = 5)$$

Si se hubiera constatado que este resultado es un redondeo inaceptable de la desviación estándar, entonces se podría dar como:

$$\bar{x} \pm s = 10.10_2 \pm 0.01_3 \quad (n = 5)$$

donde el uso de los subíndices nos indica que el dígito sólo se proporciona para evitar la pérdida de información. El lector podría decidir si fue útil o no.

De igual manera, cuando se calculan los límites de confianza [véase la Ec. (2.9)], no es necesario dar el resultado de  $ts/\sqrt{n}$  con más de dos cifras significativas. El valor de  $\bar{x}$  debería darse, en este caso, con el correspondiente número de cifras decimales.

El número de cifras significativas citado se utiliza a menudo en lugar de una estimación específica para señalar la precisión de un resultado. Por ejemplo, se considera que en el número 0.104 6*M* las tres primeras cifras decimales son ciertas, pero existen dudas acerca de la cuarta. Sin embargo, puesto que existe incertidumbre en la última cifra, podría ser cualquiera desde 0.000 05 a 0.000 5; este método proporciona una estimación pobre de la precisión y no es recomendable. A veces se recalca la incertidumbre en la última cifra al utilizar los formatos 0.104(6)*M*, o más común 0.104*M*, pero sigue siendo preferible dar una estimación específica de la precisión tal como la desviación estándar.

Un problema que puede surgir es si un 5 debería redondearse por encima o por debajo. Por ejemplo, si 9.65 se redondea a una sola cifra decimal, ¿podría convertirse en 9.6 ó 9.7? Es evidente que los resultados tendrán sesgo si un 5 se redondea siempre hacia el valor superior; este sesgo puede evitarse redondeando el 5 al número *par* más próximo, dando, en este caso, 9.6. De la misma forma 4.75 se redondea a 4.8.

Cuando hay que utilizar varias cantidades para calcular un resultado final (véase la sección 2.7), éstas no deben redondearse demasiado porque se producirá una pérdida de precisión innecesaria. Una buena regla es dejar un dígito detrás de la última cifra significativa y dejar el posterior redondeo hasta que se llegue al resultado final. La misma regla se aplica cuando se usan la media y la desviación estándar en pruebas estadísticas tales como las pruebas  $F$  y  $t$  (véase capítulo 3): se debería utilizar en los cálculos los valores no redondeados de  $\bar{x}$  y  $s$ .

## 2.6 OTROS USOS DE LOS LÍMITES DE CONFIANZA

Los límites de confianza se pueden utilizar como una prueba para detectar errores sistemáticos, como se muestra en el siguiente ejemplo.

*Ejemplo.* Se comprueba la escala de absorbancia de un espectrómetro a una longitud de onda concreta usando una solución estándar con una absorbancia de 0.470. Diez mediciones de absorbancia con el espectrómetro dieron  $\bar{x} = 0.461$  y  $s = 0.003$ . Encuentre el intervalo de confianza al 95% de la absorbancia media y decida si se encuentra presente un error sistemático.

Los límites de confianza al 95% de las absorbancias medidas por el espectrómetro son [Ec. (2.9)]:

$$\begin{aligned}\mu &= \bar{x} \pm t(s/\sqrt{n}) \\ &= 0.461 \pm 2.26 \times 0.003/\sqrt{10} \\ &= 0.461 \pm 0.002\end{aligned}$$

(El valor de  $t$  se obtuvo a partir de la tabla A.1 que se encuentra en los apéndices, la cual es una versión más completa de la tabla 2.4.)

Ya que este intervalo de confianza no incluye la absorbancia conocida de 0.470, es probable que exista un error sistemático. En el próximo capítulo se tratará otra aproximación al mismo problema.

También se pueden usar los límites de confianza cuando se realizan mediciones a cada uno de una serie de ejemplares. Por ejemplo, si se requiere el peso medio de una pastilla en un gran lote, se consumiría demasiado tiempo en pesar cada una. De igual manera, si se usa un método analítico destructivo, como la espectrometría de absorción atómica, para analizar el lote y establecer el contenido medio, sería imposible examinar cada tableta. En ambos casos, podría tomarse una muestra del lote (que en tales circunstancias forma la población) y a partir de la media y la desviación estándar de la muestra se podría encontrar un intervalo de confianza para el valor medio de la cantidad medida.

## 2.7 PROPAGACIÓN DE ERRORES ALEATORIOS

En el trabajo experimental, se calcula con frecuencia la cantidad que se va a determinar a partir de una combinación de cantidades observables. Ya hemos visto, por ejemplo, que incluso una operación relativamente simple como un análisis volumétrico, implica varias

etapas, y cada una de ellas se encuentra sujeta a errores (véase el capítulo 1). El cálculo final puede conllevar operaciones tales como sumas, restas, multiplicaciones o divisiones de dos o más cantidades, o elevar alguna cantidad a una potencia.

Es muy importante señalar que los procedimientos utilizados para combinar errores aleatorios y errores sistemáticos son completamente distintos. Esto se debe a que algunos errores aleatorios se compensan entre sí, mientras que cada error sistemático ocurre en un sentido definido y conocido. Por ejemplo, si el resultado final de un experimento,  $x$ , está dado por  $x = a + b$ ; si  $a$  y  $b$  tienen un error sistemático de  $+1$ , es evidente que el error sistemático de  $x$  es de  $+2$ . Sin embargo, si  $a$  y  $b$  tienen un error aleatorio de  $\pm 1$ , el error aleatorio de  $x$  no es 2: esto se debe a que habrá ocasiones en que el error aleatorio en  $a$  será positivo mientras que en otras en  $b$  será negativo (o viceversa).

En esta sección sólo se considera la propagación de errores aleatorios (los errores sistemáticos se consideran en la sección 2.8). Si se conoce la precisión de cada observación, entonces se pueden usar reglas matemáticas sencillas para estimar la del resultado final. Estas reglas se resumen a continuación.

### (i) Combinaciones lineales

En este caso el valor final,  $y$ , se calcula a partir de una combinación lineal de cantidades medidas  $a$ ,  $b$ ,  $c$ , etc., por:

$$y = k + k_a a + k_b b + k_c c + \dots \quad (2.10)$$

donde  $k$ ,  $k_a$ ,  $k_b$ ,  $k_c$ , etc., son constantes. La varianza (definida como el cuadrado de desviación estándar) tiene la propiedad de que la varianza de una suma o diferencia de cantidades *independientes* es igual a la suma de sus varianzas. Se puede demostrar que si  $\sigma_a$ ,  $\sigma_b$ ,  $\sigma_c$ , etc., son las desviaciones estándar de  $a$ ,  $b$ ,  $c$ , etc., la desviación estándar de  $y$ ,  $\sigma_y$ , está dada por:

$$\sigma_y = \sqrt{(k_a \sigma_a)^2 + (k_b \sigma_b)^2 + (k_c \sigma_c)^2 + \dots} \quad (2.11)$$

*Ejemplo.* En una titulación, la lectura inicial en una bureta es de 3.51 ml y la lectura final es de 15.67 ml, ambas con una desviación estándar de 0.02 ml. ¿Cuál es el volumen del titulante utilizando y cuál es su desviación estándar?

$$\begin{aligned}\text{Volumen utilizado} &= 15.67 - 3.51 = 12.16 \text{ ml} \\ \text{Desviación estándar} &= \sqrt{(0.02)^2 + (0.02)^2} = 0.028 \text{ ml}.\end{aligned}$$

Este ejemplo aclara el aspecto importante de que la desviación estándar del resultado final es mayor que la de las lecturas individuales de la bureta, incluso aunque el volumen utilizado se calcule a partir de una diferencia, pero es menor que la suma de las desviaciones estándar.

### (ii) Expresiones multiplicativas

Si  $y$  se calcula a partir de una expresión del tipo:

$$y = kab/cd \quad (2.12)$$



(donde  $a$ ,  $b$ ,  $c$  y  $d$  son cantidades medidas independientemente y  $k$  es una constante) entonces existe una relación entre los cuadrados de las desviaciones estándares *relativas*:

$$\frac{\sigma_y}{y} = \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 + \left(\frac{\sigma_c}{c}\right)^2 + \left(\frac{\sigma_d}{d}\right)^2} \quad (2.13)$$

*Ejemplo.* El rendimiento cuántico de fluorescencia,  $\phi$ , se calcula a partir de la expresión:

$$\phi = I_f / k c l I_0 \epsilon$$

donde los parámetros implicados se definen a continuación, con una estimación de sus desviaciones típicas entre paréntesis:

$I_0$  = intensidad de la luz incidente (0.5%)

$I_f$  = intensidad de luz fluorescente (2%)

$\epsilon$  = absortividad molar (1%)

$c$  = concentración (0.2%)

$l$  = paso óptico (0.2%)

$k$  es una constante del instrumento.

De la ecuación (2.13), la desviación estándar relativa de  $\phi$  está dada por:

$$\begin{aligned} \text{d.e.r.} &= \sqrt{(2^2 + 0.2^2 + 0.2^2 + 0.5^2 + 1^2)} \\ &= \sqrt{(4 + 0.04 + 0.04 + 0.25 + 1)} \\ &= \sqrt{5.33} \\ &= 2.3\% \end{aligned}$$

Puede observarse que la desviación estándar relativa del resultado final no es mucho mayor que la mayor de todas las desviaciones estándar relativas usadas para calcularlo (es decir, 2% para  $I_f$ ). Esto es, fundamentalmente, una consecuencia de elevar al cuadrado las desviaciones estándar relativas y explica una cuestión de importancia general: cualquier esfuerzo por mejorar la precisión de un experimento debe dirigirse a mejorar la precisión de los valores menos precisos. Como corolario de esto, no vale la pena esforzarse por incrementar la precisión de los valores más precisos. Esto no significa que los errores pequeños no sean importantes: éstos, en muchas etapas de un experimento, tales como las del análisis volumétrico analizados con detalle en el capítulo 1, producirán un error apreciable en el resultado final.

Es importante resaltar que cuando una cantidad se eleva a una potencia, p.ej.  $b^3$ , entonces no se calcula el error como para una multiplicación, es decir,  $b \times b \times b$ , debido a que las cantidades implicadas no son independientes. Si la relación es:

$$y = b^n \quad (2.14)$$

entonces las desviaciones típicas de  $y$  y  $b$  están relacionadas por:

$$\frac{\sigma_y}{y} = \left| \frac{n \sigma_b}{b} \right| \quad (2.15)$$

(El símbolo del módulo  $||$  significa que la magnitud de la cantidad encerrada en él se toma sin tener en cuenta el signo, por ejemplo:  $|-2| = 2$ .)

### (iii) Otras funciones

Si  $y$  es una función general de  $x$ :

$$y = f(x) \quad (2.16)$$

entonces las desviaciones estándar de  $x$  y  $y$  están relacionadas por:

$$\sigma_y = \left| \sigma_x \frac{dy}{dx} \right| \quad (2.17)$$

*Ejemplo:* la absorbancia,  $A$ , de una solución está dada por:

$$A = -\log(T) \quad (2.18)$$

donde  $T$  es la transmitancia. Si el valor medio de  $T$  es 0.501 con una desviación estándar de 0.001, calcule  $A$  y su desviación estándar.

Tenemos:

$$A = -\log 0.501 = 0.300$$

También:

$$dA/dT = -(\log e)/T = -0.434/T$$

de manera que de la ecuación (2.17)

$$\begin{aligned} \sigma_A &= |\sigma_T (-\log e/T)| \\ &= |0.001 \times (-0.434/0.501)| = 0.00087 \end{aligned}$$

Es importante observar que para este método experimental muy utilizado se pueden encontrar las condiciones para que sea mínima la desviación estándar relativa. La desviación estándar relativa de  $A$  está dada por

$$\text{e.e.m. de } A = 100 \sigma_A / A = \frac{100 \sigma_T \log e}{T \log T} \quad (2.19)$$

La derivada de esta expresión con respecto a  $T$  muestra que el coeficiente de variación de  $A$  es mínimo cuando  $T = 1/e = 0.368$ .

*Ejemplo:* La ecuación de Nernst, aplicada en el análisis potenciométrico, es

$$E = E^\circ + (RT/nF) \ln c$$

donde  $E$  es el potencial de electrodo que se mide;  $E^\circ$  es el potencial de electrodo estándar del ion que se determina;  $T$  es la temperatura absoluta;  $n$  es el número de electrones

que participan en la semicélula correspondiente;  $c$  es la concentración del ion;  $R$  es la constante de los gases, y  $F$  es la constante de Faraday. Obtenga una expresión para la desviación estándar relativa de la concentración, suponiendo que  $T = 298 \text{ K}$  y que no tiene error. Calcule el valor de la desviación estándar relativa si  $n = 1$  y la d.e. de  $E$  es  $0.001 \text{ V}$ .

Si se reescribe la ecuación de Nernst y se sustituyen los valores numéricos de  $T$ ,  $R$  y  $F$  resulta:

$$c = \exp\{40n(E - E^\circ)\}$$

Derivando con respecto a  $E$ :

$$dc/dE = 40n \exp\{40n(E - E^\circ)\}$$

de manera que de la ecuación (2.17)

$$\begin{aligned}\sigma_c &= 40n \sigma_E \exp\{40n(E - E^\circ)\} \\ &= 40nc \sigma_E\end{aligned}$$

Así, la desviación estándar relativa de  $c$  es

$$100\sigma_c/c = 100 \times 40n \sigma_E \%$$

Cuando  $n = 1$  y  $\sigma_E$  es  $0.001 \text{ V}$ , tenemos

$$\text{d.e.r. de } c = 100 \times 0.001 \times 40 = 4\%$$

En contraste con el ejemplo anterior, la desviación estándar relativa es constante.

## 2.8 PROPAGACIÓN DE ERRORES SISTEMÁTICOS

Las reglas de combinación de errores sistemáticos pueden dividirse también en tres grupos:

### (i) Combinaciones lineales

Si  $y$  se calcula a partir de cantidades medidas usando la ecuación (2.10), y los errores sistemáticos de  $a$ ,  $b$ ,  $c$ , etc., son  $\Delta a$ ,  $\Delta b$  y  $\Delta c$ , etc., entonces el error sistemático de  $y$ ,  $\Delta y$ , se calcula a partir de:

$$\Delta y = k_a \Delta a + k_b \Delta b + k_c \Delta c + \dots \quad (2.20)$$

Recuerde que los errores sistemáticos son tanto positivos como negativos y que estos signos *deben* incluirse en el cálculo de  $\Delta y$ .

El error sistemático total puede ser a veces cero. Por ejemplo, suponga que se usa una balanza con error sistemático de  $-0.01 \text{ g}$  para pesadas utilizadas en la preparación de una solución estándar. Puesto que el peso de soluto utilizado se calcula a partir de una

diferencia entre dos pesadas, se eliminan los errores sistemáticos. Procedimientos como éste, considerados en forma cuidadosa, pueden minimizar a menudo los errores sistemáticos, como se describió en el capítulo 1. Sin embargo, es preciso señalar que esto se aplica sólo a una balanza electrónica con una única pesa de referencia interna; cuando se usa una balanza "pasada de moda" con una serie de pesas individuales, éstas deben calibrarse, y hacerse las correcciones correspondientes (¡y todavía tendrán un error de calibrado aleatorio!).

### (ii) Expresiones multiplicativas

Si  $y$  se calcula a partir de cantidades mediciones utilizando la ecuación (2.12), entonces se utilizan errores sistemáticos *relativos*:

$$(\Delta y/y) = (\Delta a/a) + |\Delta b/b| + (\Delta c/c) + (\Delta d/d) \quad (2.21)$$

Cuando se eleva una cantidad a alguna potencia, entonces se emplea la ecuación (2.15), se omite el símbolo del módulo y se sustituyen las desviaciones estándar por errores sistemáticos.

### (iii) Otras funciones

La ecuación correspondiente es idéntica a la ecuación (2.17), pero se omite el símbolo del módulo y se sustituyen las desviaciones estándares por errores sistemáticos.

## BIBLIOGRAFÍA

Davies, O. L. y Goldsmith, P. L., *Statistical Methods in Research and Production*. Londres, Longmans, 1982. Se trata con más detalle el tema de este capítulo.

Móritz, P., Capítulo 1 en *Comprehensive Analytical Chemistry Vol. XI*, Svehla, G. Ed., Amsterdam, Elsevier, 1981. En este artículo se profundiza en la aplicación de la teoría de errores a los métodos analíticos.

Skoog, D. A. y West, D. M., *Fundamentals of Analytical Chemistry*, 4a ed., Nueva York, Holt Saunders, 1982. Se describe el uso de la estadística en la evaluación de datos analíticos.

Topping, J., *Errors of Observation and their Treatment*, Londres, Chapman & Hall, 1962. Se trata de manera más completa la teoría de errores y se estudia la teoría de que los errores se distribuyen normalmente.

## EJERCICIOS

1. Para investigar la reproducibilidad de un método para la determinación de selenio en alimentos, se realizaron nueve mediciones sobre un lote de arroz tostado, con los siguientes resultados:

Muestras	Selenio, $\mu\text{g/g}$
1	0.07
2	0.07
3	0.08
4	0.07

Muestras	Selenio, $\mu\text{g/g}$
5	0.07
6	0.08
7	0.08
8	0.09
9	0.08

(Moreno Domínguez, T., García Moreno, C. y Mariné-Font, A., *Analyst*, 1983, **108**, 505).

Calcule la media, la desviación estándar y la desviación estándar relativa de estos resultados.

2. Siete mediciones de pH de una solución reguladora proporcionaron los siguientes resultados:

5.12, 5.20, 5.15, 5.17, 5.16, 5.19, 5.15.

Calcule el verdadero pH al nivel de confianza del (i) 95% y (ii) 99%. (Suponga que no existen errores sistemáticos.)

3. Diez análisis de la concentración de albúmina dieron una medida de 20.92 g/l, y una desviación estándar de 0.45 g/l. (Foote, J. W. y Delves, H. T., *Analyst*, 1983, **108**, 492.)  
Calcule los límites de confianza de la medida al 95%.

4. Se midió la concentración de plomo en el fluido sanguíneo para una muestra de 50 niños de un gran colegio próximo a una calle concurrida. La medida muestral fue 10.12 ng/ml y la desviación estándar fue 0.64 ng/ml. Calcule el intervalo de confianza al 95% para la concentración media de plomo de todos los niños de la escuela.

¿Qué tamaño debería tener la muestra para reducir la longitud del intervalo de confianza a 0.2 ng/ml (es decir  $\pm 0.1$  ng/ml)?

5. En una investigación de la exactitud y precisión de un método para la determinación de arprinocida en comestibles preparados, se realizaron seis determinaciones repetidas de una formulación preparada que contiene 10.2% de arprinocida. Los resultados fueron:

10.4, 10.4, 10.6, 10.3, 10.5, 10.5%.

(Stong, J. D. y Fink, D. W., *Analyst*, 1982, **107**, 113).

Calcule la media, la desviación estándar y los límites de confianza para la media al 95% y 99% ¿Se encuentra el valor de 10.2% dentro de los límites de confianza al nivel del (i) 95% y (ii) del 99%?

6. La medida de la concentración de haptoglobina en suero sanguíneo tomada de una muestra al azar de ocho adultos sanos dio los siguientes resultados:

1.82, 3.32, 1.07, 1.27, 0.49, 3.79, 0.15, 1.98 g/l

Calcule la media y la desviación estándar de estos resultados.

Suponiendo que la distribución de la haptoglobina es log-normal en la población como conjunto, calcule el intervalo de confianza al 95% para la concentración media de haptoglobina para el conjunto de la población.

7. Diez mediciones de la relación entre las áreas de dos picos en un experimento de cromatografía líquida dieron los siguiente valores:

0.2911, 0.2898, 0.2923, 0.3019, 0.2997, 0.2961, 0.2986, 0.2902, 0.2882.

(Jonvel, P. y Anderman, G., *Analyst*, 1983, **108**, 411).

Calcule la media, la desviación estándar y los límites de confianza de la media al 99%.

8. Se utilizó una disolución de 0.1M de ácido para valorar 10 ml de una solución de álcali de 0.1M y se registraron los siguientes volúmenes de ácido:

9.88, 10.18, 10.23, 10.39, 10.25 ml.

Calcule los límites de confianza de la media al 95% y utilícelos para decir si existe alguna evidencia de error sistemático.

9. En este problema se consideran los errores aleatorios involucrados en la preparación de una disolución estándar. Se preparó un volumen de 250 ml de una disolución 0.05M de un reactivo de peso molecular 40, por diferencia de pesada. La desviación estándar de cada pesada fue de 0.0001 g; ¿cuál fue la desviación estándar y la desviación estándar relativa del peso del reactivo utilizado? La desviación estándar del volumen del disolvente utilizado fue de 0.05 ml. Expresé esto como una desviación estándar relativa. De aquí calcule la desviación estándar relativa de molaridad de la solución.

Repita el cálculo para un reactivo de peso molecular 392.

10. El producto de solubilidad del sulfato de bario es  $1.3 \times 10^{-10}$ , con una desviación estándar de  $0.1 \times 10^{-10}$ . Calcule la desviación estándar de la solubilidad calculada del sulfato de bario en agua.

## Pruebas de significación

### 3.1 INTRODUCCIÓN

Una de las propiedades más importantes de un método analítico es que se encuentre libre de errores sistemáticos, es decir, el valor dado para la cantidad de analito debería ser el valor *verdadero*. Esta propiedad se puede comprobar al aplicar el método a una muestra estándar que contenga una cantidad conocida de analito (véase el Cap. 1). Sin embargo, como se vio en el capítulo anterior, los errores aleatorios no permiten que la cantidad medida sea *exactamente* igual a la cantidad conocida incluso aunque no hubiera error sistemático. Para decidir si la diferencia entre la cantidad medida y la cantidad conocida se puede justificar por estos errores aleatorios, puede aplicarse una prueba estadística que se denomina **prueba de significación**. Como su nombre lo indica, esta aproximación prueba si son significativas las diferencias entre los dos resultados, o si se pueden justificar sólo por variaciones aleatorias. Las pruebas de significación se han utilizado en forma amplia en la evaluación de los resultados experimentales. En este capítulo se consideran varias pruebas que son útiles particularmente para los químicos analíticos.

### 3.2 COMPARACIÓN DE UNA MEDIA EXPERIMENTAL CON UN VALOR CONOCIDO

Al realizar una prueba de significación comprobamos la veracidad de una hipótesis denominada **hipótesis nula**. Como observamos en el párrafo anterior, adoptamos como hipótesis nula aquella mediante la cual un método *no* se encuentra sujeto a errores sistemáticos. El término *nulo* se utiliza para indicar que *no* hay más diferencia, entre lo observado y el valor conocido, que la que puede atribuirse a la variación aleatoria. Suponiendo que esta hipótesis nula es verdadera, se puede utilizar la teoría estadística para calcular la probabilidad (es decir, la posibilidad) de que la diferencia observada entre la media muestral,  $\bar{x}$ , y el verdadero valor,  $\mu$ , se deba solamente a un error aleatorio. Cuando más pequeña sea la probabilidad de que la diferencia observada ocurra por azar, menos

probable será que la hipótesis nula sea verdadera. Por lo regular, la hipótesis nula se rechaza cuando la probabilidad de que la diferencia observada ocurra por azar es *menor* que 1 de cada 20 veces (es decir, 0.05 ó 5%), y en este caso se dice que la diferencia es **significativa al nivel de 0.05 (o 5%)**. Si utilizamos este nivel de significación rechazamos en promedio la hipótesis nula, *cuando sea de hecho verdadera*, una de cada veinte veces. Para estar más seguros de que se toma la decisión adecuada, se utiliza un nivel de significación más pequeño, por lo regular 0.01 ó 0.001 (1% o 0.1%). El nivel de significación se indica escribiendo, por ejemplo,  $P$  (es decir, la probabilidad) = 0.05. Es importante tener en cuenta que si se acepta la hipótesis nula no significa que hayamos *probado* que sea verdadera, sólo que no hemos demostrado que sea falsa. Más adelante en este capítulo se analizará la probabilidad de aceptar una hipótesis *nula cuando sea de hecho falsa*.

Para decidir si la diferencia entre  $\mu$  y  $\bar{x}$  es significativa, la ecuación (2.9):

$$\mu = \bar{x} \pm (ts/\sqrt{n})$$

(donde  $n$  es el tamaño muestral) se escribe de otra forma:

$$t = (\bar{x} - \mu)\sqrt{n/s} \quad (3.1)$$

y se calcula un valor de  $t$  sustituyendo los resultados experimentales en esta ecuación. Si  $|t|$  (es decir, el valor de  $t$  sin considerar el signo) es *mayor* que un cierto **valor crítico**, entonces se rechaza la hipótesis nula. El valor crítico de  $|t|$  para un nivel de significación concreto se encuentra en la tabla A.1. Por ejemplo, para un tamaño de muestra de 10 (es decir, 9 grados de libertad) y un nivel de significación de 0.01, el valor crítico de  $|t|$  es 3.25.

*Ejemplo.* En un método para determinar mercurio por la técnica de absorción atómica de vapor frío, se obtuvieron los siguientes valores para un material de referencia que contiene 38.9% de mercurio:

38.9, 37.4, 37.1%

(Hou, P. K., Lau, O. W. y Wong, M. C., *Analyst*, 1983, **108**, 64.)

¿Hay alguna evidencia de error sistemático?

La media de estos valores es de 37.8% y la desviación estándar de 0.964%. Si se adopta la hipótesis nula de que no hay error sistemático, es decir,  $\mu = 38.9\%$  y se utiliza la ecuación (3.1) resulta

$$|t| = |(37.8 - 38.9) \times \sqrt{3} / 0.964| = 1.98$$

De la tabla A.1, para 2 grados de libertad, el valor crítico de  $|t|$  es 4.3 ( $P = 0.05$ ). Ya que el valor observado de  $|t|$  es menor que el valor crítico, no se rechaza la hipótesis nula: no hay evidencia de error sistemático. Hay que señalar nuevamente que esto no significa que no *existan* errores sistemáticos, sino que no se ha podido constatar su existencia.

### 3.3 COMPARACIÓN DE LAS MEDIAS DE DOS MUESTRAS

Otra forma por la que los resultados de un método analítico nuevo pueden comprobarse es por comparación de los obtenidos utilizando un segundo método (quizás un método de referencia). En este caso tenemos dos medias muestrales  $\bar{x}_1$  y  $\bar{x}_2$ . Si tomamos como hipótesis nula que los dos métodos dan el mismo resultado, necesitamos comprobar si  $(\bar{x}_1 - \bar{x}_2)$  difiere en forma significativa de cero. Si las dos muestras tienen desviaciones estándar que no sean significativamente diferentes (véase sección 3.6, en donde se presenta un método para probar esta suposición), se puede calcular una estimación **conjunta** de la desviación estándar a partir de la dos desviaciones estándar individuales  $s_1$  y  $s_2$  utilizando la ecuación:

$$s^2 = \{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2\}/(n_1 + n_2 - 2) \quad (3.2)$$

Se puede demostrar que entonces  $t$  está dado por:

$$t = (\bar{x}_1 - \bar{x}_2)/s\sqrt{(1/n_1 + 1/n_2)} \quad (3.3)$$

Donde  $t$  tiene  $n_1 + n_2 - 2$  grados de libertad.

*Ejemplo.* En una comparación de dos métodos para la determinación de boro en muestras de plantas se obtuvieron los siguientes resultados ( $\mu\text{g/g}$ ):

1 método espectrofotométrico: media = 28.0; desviación estándar = 0.3

2 método fluorimétrico: media = 26.25; desviación estándar = 0.23

Para cada método se realizaron diez determinaciones.

(Anarez, J., Bonilla, A. y Vidal, J. C., *Analyst*, 1983, **108**, 368.)

¿Estos dos métodos dan resultados cuyas medidas difieren significativamente?

La hipótesis nula adoptada es la que la media de los resultados dada por ambos métodos es la misma. De la ecuación (3.2), el valor de la desviación estándar conjunta es:

$$\begin{aligned} s^2 &= (9 \times 0.3^2 + 9 \times 0.23^2)/18 \\ &= 0.0715 \\ s &= 0.267 \end{aligned}$$

De la ecuación (3.3):

$$\begin{aligned} t &= (28.0 - 26.25)/0.267\sqrt{(1/10 + 1/10)} \\ &= 14.7 \end{aligned}$$

Hay 18 grados de libertad, por lo tanto (tabla A.1) el valor crítico de  $|t|$  ( $P = 0.05$ ) es 2.1; puesto que el valor experimental de  $|t|$  es más grande que éste, de la diferencia entre los resultados es significativa al nivel del 5% y se rechaza la hipótesis nula. De hecho, ya que el valor crítico de  $|t|$  para  $P = 0.001$  es aproximadamente de 3.9, la diferencia es significativa incluso al nivel del 0.1%. En otras palabras, si la hipótesis nula es

verdadera, la probabilidad de que haya una gran diferencia debida al azar es menor que 1 entre 1000.

El siguiente ejemplo, es otra aplicación de esta prueba, en donde se emplea para decidir si un cambio en las condiciones de un experimento afecta el resultado.

*Ejemplo.* En una serie de experimentos para la determinación de estaño en productos alimenticios, las muestras se llevaron al punto de ebullición con HCl a reflujo durante diferentes tiempos. Los resultados fueron:

Tiempo de reflujo (min)	Estaño encontrado (mg/kg)
30	55, 57, 59, 56, 56, 59
70	57, 55, 58, 59, 59, 59

(Analytical Methods Committee, *Analyst*, 1983, **108**, 109.)

¿Es diferente la cantidad media de estaño encontrada para los dos tiempos de ebullición?

La media y la varianza (cuadrado de la desviación estándar) para los dos tiempos son:

30 min	$\bar{x}_1 = 57.0$	$s_1^2 = 2.80$
75 min	$\bar{x}_2 = 57.8$	$s_2^2 = 2.57$

La hipótesis nula adoptada es que el tiempo de ebullición no influye en la cantidad de estaño encontrada. De la ecuación (3.2), el valor para la varianza está dado por:

$$\begin{aligned} s^2 &= (5 \times 2.80 + 5 \times 2.57)/10 \\ &= 26.85 \\ s &= 1.64 \end{aligned}$$

$t$  se calcula a partir de la ecuación (3.3):

$$\begin{aligned} t &= (57.0 - 57.8)/1.64\sqrt{(1/6 + 1/6)} \\ &= -0.84 \end{aligned}$$

Hay 10 grados de libertad, de manera que el valor crítico de  $|t|$  es 2.23 ( $P = 0.05$ ). El valor observado de  $|t|$  es menor que el valor crítico, por lo tanto, se acepta la hipótesis nula: no hay pruebas de que el periodo de ebullición afecte la tasa de recuperación.

Si no es válido suponer que las desviaciones estándar de las poblaciones son iguales, un método aproximado consiste en modificar la ecuación (3.3) por:

$$t = (\bar{x}_1 - \bar{x}_2)/\sqrt{(s_1^2/n_1 + s_2^2/n_2)} \quad (3.4)$$

y calcular el número de grados de libertad a partir de:

$$\text{grados de libertad} = \left\{ \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1 + 1} + \frac{(s_2^2/n_2)^2}{n_2 + 1}} \right\} - 2 \quad (3.5)$$

el resultado se redondea al número entero más cercano.

*Ejemplo.* La siguiente tabla proporciona la concentración de tiol en sangre de dos grupos de voluntarios, el primer grupo es "normal" y el segundo sufre artritis reumatoide.

Concentración de tiol ( mM )	
Normal	Reumatoide
1.84	2.81
1.92	4.06
1.94	3.62
1.92	3.27
1.85	3.27
1.91	3.76
2.07	

(Banford, J. C., Brown, D. H., McConnell, A. A., McNeil, C. J., Smith, W. E., Hazelton, R. A. y Sturrock, R. D., *Analyst*, 1983, **107**, 195.)

De los datos anteriores tenemos que:

$$\begin{aligned} n_1 &= 7, & \bar{x}_1 &= 1.921, & s_1 &= 0.076 \\ n_2 &= 6, & \bar{x}_2 &= 3.465, & s_2 &= 0.440 \end{aligned}$$

De nueva cuenta, la hipótesis nula es que la concentración media de tiol es la misma para los dos grupos. Si se sustituye en la ecuación (3.4) resulta  $t = 8.5$  y de la ecuación (3.5) se obtienen 5 grados de libertad. El valor crítico de  $|t|$  ( $P = 0.001$ ) es 4.03, de manera que se rechaza la hipótesis nula: la concentración media de tiol es diferente para los dos grupos.

### 3.4 LA PRUEBA $t$ POR PAREJAS

Con frecuencia ocurre que deben compararse dos métodos de análisis por medio del estudio de muestras que contienen, de manera sustancial, diferentes cantidades de analito, como se ilustra en el siguiente ejemplo:

*Ejemplo.* La siguiente tabla proporciona la concentración de plomo ( $\mu\text{g/l}$ ) determinado por dos métodos diferentes para cada una de las cuatro muestras:

Muestras	Oxidación húmeda	Extracción directa
1	71	76
2	61	68
3	50	48
4	60	57

(Smith, B. M. y Griffiths, M. B., *Analyst*, 1982, **107**, 253.)

¿Los dos métodos proporcionaron valores para las concentraciones medias de plomo que difieran significativamente?

La prueba para comparar dos medias (véase la Sec. 3.3) no es apropiada en este caso, ya que cualquier variación debida al método se verá afectada por causa de las diferencias entre las muestras comparadas. Se resuelve esta dificultad al observar la diferencia entre cada par de resultados dados por dos métodos. Si adoptamos la hipótesis nula de que no existen diferencias significativas en las concentraciones dadas por los dos métodos, se puede probar si la media de las diferencias difiere significativamente de cero.

Para los pares de valores anteriores las diferencias son -5, -7, 2 y 3. La media de las diferencias,  $\bar{x}_d$ , es -1.75 y la desviación estándar de las diferencias,  $s_d$ , es 4.99. Puesto que  $\mu_d = 0$ , la ecuación (3.1) para calcular  $t$  se transforma en:

$$t = \bar{x}_d \sqrt{n}/s_d \quad (3.6)$$

donde  $t$  tiene  $(n-1)$  grados de libertad. Si se sustituye en la ecuación (3.6) resulta un valor de  $t = -0.7$ . El valor crítico de  $|t|$  es 3.18 ( $P = 0.05$ ) y puesto que el valor calculado de  $|t|$  es menor que éste, se acepta la hipótesis: el método no da valores significativamente diferentes para las concentraciones medias de plomo.

Hay varias circunstancias por las cuales puede ser necesario o deseable diseñar un experimento de manera que se analice cada muestra por cada uno de los dos métodos y, de este modo, los resultados surjan emparejados en forma natural. Algunos ejemplos son:

- la cantidad de muestra disponible que se va a examinar sólo es suficiente para una determinación por cada método;
- los métodos se van a comparar utilizando una gran variedad de muestras de diferente procedencia y posiblemente con concentraciones muy distintas (véase el próximo párrafo);
- las muestras que se van a examinar pueden presentarse durante un periodo extenso por lo que es necesario eliminar los efectos de las condiciones ambientales tales como temperatura, presión, etcétera.

Como es común que los métodos analíticos se apliquen en un amplio intervalo de concentraciones, a menudo se compara un método nuevo con uno estándar mediante el análisis de muestras en las que las características del analito puedan variar en varias potencias de diez. En este caso resulta inadecuado utilizar la prueba  $t$  por parejas, ya que su validez se apoya en la hipótesis de que cualquier error, ya sea aleatorio o sistemático,

es independiente de la concentración. En intervalos de concentración amplios dicha hipótesis no tiene por que ser del todo cierta. El método estadístico preferido en estos casos es la regresión lineal: más adelante se describe su aplicación (véase la Sec. 5.9).

### 3.5 LAS PRUEBAS DE UNA Y DOS COLAS

Los métodos analizados hasta el momento en este capítulo se refieren a probar la diferencia entre dos medias *en cualquier dirección*. Por ejemplo, el método descrito en la sección 3.2 prueba si existe diferencia significativa entre el resultado experimental y el valor conocido para el material de referencia, sin tener en cuenta el signo de la diferencia. En muchas situaciones de este tipo, el analista no tiene una idea preconcebida, *previa a las mediciones experimentales*, con relación a si la diferencia entre la media experimental y el valor de referencia será positiva o negativa.

De esta manera, es necesario contar con una prueba que cubra esta posibilidad: ésta se denomina prueba de **dos colas** (o **bilateral**). Sin embargo, en algunos casos puede ser apropiado utilizar un tipo diferente de prueba. Por ejemplo, en un experimento se espera *incrementar* la velocidad de reacción añadiendo un catalizador; es evidente que *antes de empezar el experimento* sólo se tiene interés en que la velocidad nueva sea *mayor* que la anterior y, por lo tanto, sólo es necesario probar un incremento. Este tipo de prueba se llama de **una cola** (o **unilateral**). Para un valor dado de  $n$  y un nivel de probabilidad concreto, el valor crítico para una prueba unilateral difiere de una bilateral. En una prueba unilateral, para un incremento, el valor crítico de  $t$  (en lugar de  $|t|$ ) para  $P = 0.05$  es aquel que es *superado* con una probabilidad del 5%. Ya que se supone que la distribución muestral de la media es simétrica, esta probabilidad es la mitad de la probabilidad de que sea importante en la prueba bilateral. De esta manera, para encontrar el valor adecuado para la prueba de una cola se busca en la columna de  $P = 0.10$  en la tabla A.1. En forma similar, para una prueba de una cola con  $P = 0.01$ , se emplea el valor en la columna  $P = 0.02$ . Para una prueba de una cola para un *decremento*, el valor crítico de  $t$  será de la misma magnitud pero con un signo negativo.

*Ejemplo.* Se sospecha que una valoración ácido-base tiene un error de indicador significativo y, por lo tanto, tiende a dar resultados con un error sistemático positivo (es decir, un sesgo positivo). Para comprobar esto, se utiliza una disolución de ácido exactamente 0.1M para valorar 25.00 ml de otra disolución de una base, exactamente 0.1M con los siguientes resultados (ml):

25.06 25.18 24.87 25.51 25.34 25.41

Pruebe la existencia de sesgo positivo en estos resultados.

De estos datos se tiene:

media = 25.228 ml; desviación estándar = 0.238 ml.

Si adoptamos la hipótesis nula de que no hay ningún sesgo, es decir,  $\mu = 25.00$ , y utilizamos la ecuación (3.1) resulta

$$t = (25.228 - 25.00) \times \sqrt{6} / 0.238 = 2.35$$

De la tabla A.1 el valor crítico de  $t$  para 5 grados de libertad es 2.02 ( $P = 0.05$ , prueba de una cola). Puesto que el valor observado de  $t$  es mayor que éste, se rechaza la hipótesis nula y hay evidencia de sesgo positivo.

Es interesante observar que si en el ejemplo anterior se hubiera realizado una prueba de dos colas ( $|t| = 2.57$ ), *no* se habría rechazado la hipótesis nula. La explicación de este hecho aparentemente contradictorio, reside en que la decisión de efectuar una prueba de una o dos colas depende del grado de conocimiento a priori, ya que en este caso se sospechaba o esperaba sesgo positivo. Por supuesto, resulta esencial que se tome la decisión sobre si la prueba será de una o dos colas *antes* de realizar el experimento, y no después, cuando los resultados podrían prejuzgar la elección. En general, se encontrarán más situaciones en las que se empleen pruebas de dos colas que de una cola, y se detectarán con facilidad las relativamente raras circunstancias en las que sea necesario emplear pruebas de una cola.

### 3.6 LA PRUEBA F PARA LA COMPARACIÓN DE DESVIACIONES ESTÁNDAR

Las pruebas de significación descritas hasta ahora en este capítulo se utilizan para comparar medias y, por lo tanto, para detectar errores sistemáticos. En muchos casos, también es importante comparar las desviaciones estándar, es decir, los errores aleatorios de dos conjuntos de datos. Esta comparación, como en el caso de las pruebas de medias, puede tener dos formas; o bien probar si el Método A es más preciso que el Método B (es decir, la prueba de una cola) o si los métodos A y B difieren en su precisión (o sea, la prueba de dos colas). Así, si quisiéramos probar si un método analítico nuevo es *más* preciso que uno estándar, deberíamos utilizar la prueba de una cola; si quisiéramos probar si dos desviaciones estándar difieren significativamente (por ejemplo, antes de aplicar la prueba  $t$ , véase la sección 3.3), sería adecuado una prueba de dos colas.

En la **prueba F** se considera la razón de las dos varianzas muestrales, es decir, la razón de los cuadrados de las desviaciones estándar. La cantidad calculada ( $F$ ) está dada por:

$$F = s_1^2 / s_2^2 \quad (3.7)$$

y se disponen los valores  $s_1^2$  y  $s_2^2$  en la expresión anterior, de manera que  $F$  sea siempre  $\geq 1$ . La hipótesis nula adoptada es que las poblaciones de donde se toman las muestras sean normales, y que sean iguales las varianzas de las poblaciones. Si la hipótesis nula es verdadera, entonces la razón de varianzas debería estar próxima a 1. Las diferencias respecto de 1 se deben a variaciones aleatorias, pero si la diferencia es demasiado grande no se podrá atribuir a esta causa: si el valor calculado de  $F$  excede un cierto valor (obtenido de las tablas), entonces se rechaza la hipótesis nula. Este valor crítico de  $F$  depende del tamaño de las dos muestras, del nivel de significación y del tipo de prueba realizada. Los valores para  $P = 0.05$  están dados en las tablas A.2 y A.3; el uso de las tablas se aclara en los siguientes ejemplos.



*Ejemplo.* Se comparó un método propuesto para la determinación de la demanda de oxígeno en aguas residuales con un método estándar (sales de mercurio). Se obtuvieron los siguientes resultados para una muestra de aguas residuales:

	Media (mg/l)	Desviación estándar (mg/l)
Método estándar	72	3.31
Método propuesto	72	1.51

Para cada método se realizaron 8 determinaciones.

(Ballinger, D., Lloyd, A. y Morrish, A., *Analyst*, 1982, **107**, 1047.)

¿Es la precisión del método propuesto significativamente más grande que la del método estándar?

Tenemos que decir si la varianza del método estándar es significativamente mayor que la del método propuesto.  $F$  está dado por la razón de varianzas:

$$F_{7,7} = 3.31^2 / 1.51^2 = 4.8$$

Se hicieron ocho determinaciones, por lo tanto, el número de grados de libertad es 7 en cada caso, como está indicado por los subíndices. En este caso hay que utilizar la prueba de una cola, ya que sólo nos interesa si el método propuesto es *más* preciso que el método estándar. En la tabla A.2, el número de grados de libertad del denominador está dado en la columna de la izquierda, y el número de grados de libertad del numerador, en la parte superior. El valor crítico de  $F$  es, en este caso, 3.787 ( $P = 0.05$ ). Ya que el valor calculado de  $F$  (4.8) lo excede, la varianza del método estándar es significativamente mayor que la del método propuesto a un nivel de probabilidad del 5%, es decir, el método propuesto es más preciso.

*Ejemplo.* En el ejemplo que se encuentra al principio de la sección 3.3 (determinación de boro en plantas), se supuso que las varianzas de los dos métodos no diferían significativamente. Esta hipótesis se puede probar ahora. Las desviaciones estándar eran 0.3 y 0.23 (cada una obtenida de 10 medidas de un ejemplar de una planta determinada). Calculando  $F$  de manera que sea mayor o igual que 1, tenemos:

$$F_{9,9} = 0.3^2 / 0.23^2 = 1.7$$

Sin embargo, en este caso no tenemos ninguna razón para esperar, de antemano, que la varianza de un método deba ser mayor que la del otro; por ello usaremos una prueba de dos colas. Los valores críticos que se proporcionan en la tabla A.2 son los valores que  $F$  supera con una probabilidad de 0.05, suponiendo que *deben* ser mayores que 1. En una prueba de dos colas, la razón entre una varianza y otra podría ser mayor o menor que 1, pero si  $F$  se calcula de manera que *sea* siempre mayor que 1, entonces la probabilidad de que supere el valor crítico dado en la tabla A.2 será doble. Por ello, estos valores críticos no son apropiados para una prueba de dos colas y en lugar de la

tabla A.2 se emplea la tabla A.3. De esta tabla, tomando como el número de grados de libertad tanto del numerador como del denominador igual a 9, el valor crítico obtenido para  $F$  es 4.026. El valor calculado es menor que éste, por lo tanto no hay diferencia significativa entre las dos varianzas a un nivel del 5%.

Como en el caso de la prueba  $t$ , en la prueba  $F$  se puede emplear otros niveles de significación y los valores críticos se encuentran en las tablas que aparecen en la bibliografía al final del capítulo 1. Se debe tener cuidado en la elección de la tabla correcta, dependiendo de si es una prueba de uno o dos colas: para un  $\alpha\%$  de nivel de significación, los  $\alpha\%$  puntos de la distribución  $F$  se utilizan para una prueba de una cola, mientras que para una prueba de dos colas se usan los puntos  $\frac{1}{2}\alpha\%$ .

### 3.7 VALORES ANÓMALOS

Es muy frecuente que las personas que realizan experimentos se enfrenten a la situación en que uno (o posiblemente más) de los resultados obtenidos en una serie de medidas difiera del resto de manera inexplicable. Estas medidas se denominan resultados **anómalos**. En algunos casos, se puede atribuir un resultado anómalo a un error humano. Por ejemplo, si se obtienen los siguientes resultados en una titulación:

12.12, 12.15, 12.13, 13.14, 12.12 ml

entonces es casi seguro que el cuarto valor corresponda a un error en la escritura y que debiera leerse como 12.14. Sin embargo, incluso cuando los valores que son obviamente erróneos han sido eliminados o corregidos, todavía quedan aquellos que pudieran ser valores anómalos. ¿Deberíamos conservarlos, o deberíamos buscar un valor medio para comprobar estadísticamente si deben ser rechazados o no? Por supuesto, los valores finales de la media y la desviación estándar dependerán de que hayan sido rechazados o no los valores anómalos. Puesto que el análisis sobre la precisión y la exactitud de un método depende de estos valores finales, siempre tiene que quedar claro si han sido rechazados los valores anómalos y, en tal caso, por qué.

Hasta aquí el análisis de los errores se ha basado en la hipótesis de que la distribución de medidas repetidas es normal (o aproximadamente). Una razón para la presencia de valores anómalos puede ser que la afirmación anterior no sea válida: nosotros suponemos un modelo concreto, es decir, normal, para la población a partir de la cual se extraen las medidas, y debido a que este modelo no es correcto, se obtiene un resultado sorprendente. En esta sección sólo se describirán pruebas de valores anómalos que supongan una población normal. Si no se admitiera la población normal, entonces no se debería emplear las pruebas descritas en este capítulo (a menos que las muestras sean grandes, ya que en este caso las distribuciones muestrales llegan a ser normales), sino que se debería sustituir por los métodos no paramétricos apropiados descritos en el capítulo 6. Estos últimos por lo general son insensibles a los valores anómalos y por ello se evita el problema de si se rechazan o no dichos valores.

Una forma (entre varias pruebas diferentes) de estudiar una medida sospechosa es comparar la diferencia entre ella y la medida más próxima en tamaño, con la diferencia

entre las medidas más grande y más pequeña. El cociente de estas diferencias (prescindiendo del signo) se denomina **Q de Dixon**.

$$Q = |\text{valor sospechoso} - \text{valor más cercano}| / (\text{valor más grande} - \text{valor más pequeño}) \quad (3.8)$$

Los valores críticos de  $Q$  para  $P = 0.05$  y  $P = 0.01$  se encuentran en la tabla A.4. Si el valor de  $Q$  calculado supera el valor crítico, se rechaza el valor sospechoso.

*Ejemplo.* Se obtuvieron los siguientes valores para la concentración de nitrito (mg/l) en una muestra de agua de río:

0.403, 0.410, 0.401, 0.380

La última medida es sospechosa, ¿debería ser rechazada?  
Tenemos:

$$Q = |0.380 - 0.401| / (0.410 - 0.380) = 0.021/0.03 = 0.7$$

De la tabla A.4, para un tamaño de muestra 4, el valor crítico de  $Q$  es 0.831 ( $P = 0.05$ ). Ya que el valor calculado de  $Q$  no lo supera, debe aceptarse la medida sospechosa.

En una situación ideal, habría que tomar nuevas medidas cuando aparezca un valor sospechoso, en especial si sólo se han tomado inicialmente unos pocos valores. Esto podría aclarar si debería rechazarse el valor sospechoso o no y, si se mantuviera, también reduciría su efecto sobre la media y la desviación estándar.

*Ejemplo.* Si se añaden tres muestras medidas a las dadas en el ejemplo anterior, quedaría:

0.403, 0.410, 0.401, 0.380, 0.400, 0.413, 0.411

¿se debería aún mantener el 0.380?

El valor de  $Q$  calculado es ahora:

$$Q = |0.380 - 0.400| / (0.413 - 0.380) = 0.606$$

El valor crítico de  $Q$  ( $P = 0.05$ ) para un tamaño muestral 7 es 0.570, por lo que la medida sospechosa se rechaza a un nivel de significación del 5%.

Es importante tener en cuenta que para un nivel de significación del 5% hay todavía un 5% de riesgo, o 1 de cada 20, de rechazar incorrectamente un valor sospechoso. Esto puede tener un efecto considerable a la hora de estimar la precisión del experimento. Por ejemplo, para los 7 valores de la concentración de nitrito dados anteriormente, la desviación estándar es de 0.011 mg/l, pero cuando se rechaza el valor sospechoso la desviación estándar se convierte en 0.0056 mg/l, es decir, la precisión parece haber mejorado en un factor de 2.

En el ejemplo anterior se subraya la importancia de ser cauto al rechazar valores anómalos. Cuando las medidas se repiten sólo unas pocas veces (lo que es normal en un

trabajo analítico), el rechazo de un valor origina una gran variación sobre la media y sobre la desviación estándar. En concreto, debe evitarse el hecho de tomar tres medidas y rechazar la que difiere más de las otras dos. Se puede demostrar que, en promedio, se obtiene una estimación más confiable de la media utilizando el valor que está en medio de los tres, más que utilizando la media de los dos que no fueron rechazados.

De la misma manera que se han descrito anteriormente pruebas para un valor anómalo, también existen otras para dos valores anómalos. Sin entrar en detalle, podemos encontrar dos valores sospechosos: los dos próximos, o uno a cada extremo del intervalo de valores, lo que produce una reducción en el valor calculado de  $Q$ . En el primer caso habrá una reducción en el numerador, y en el segundo un incremento en el denominador de la fracción. En cualquier caso, el efecto de la medida más extrema estará **enmascarado** en la prueba  $Q$  por la presencia del otro valor anómalo posible. Éste se puede comprobar al considerar los siguientes valores:

2.1, 2.0, 2.1, 2.3, 2.9, 2.3, 3.1, 2.2, 2.0, 2.3

Obviamente 2.9 y 3.1 son los valores sospechosos, pero el valor calculado de  $Q$  es:

$$Q = (3.1 - 2.9) / (3.1 - 2.0) = 0.18$$

un valor que no es significativo ( $P = 0.05$ ). En estos casos, lo apropiado es una prueba para un par de valores anómalos; las referencias de éstos aparecen en la bibliografía al final del capítulo.

Los valores anómalos aparecen también en otros cálculos estadísticos, como en los métodos de regresión lineal que se describirán en el capítulo 5. Estos casos se analizan también en textos avanzados.

### 3.8 ANÁLISIS DE LA VARIANZA

En la sección 3.3 se describió un método para comparar dos medias y probar si diferían en forma significativa. En el trabajo analítico se presentan a menudo comparaciones en las que intervienen más de dos medias. Algunas situaciones de este tipo son: comparar la concentración media de proteína en una solución para muestras almacenadas en condiciones diferentes; comparar los resultados medios obtenidos de la concentración de un analito utilizando diferentes métodos; comparar la media de los resultados en una valoración obtenidos por diferentes operadores que usan los mismos aparatos. En todos estos ejemplos hay dos posibles fuentes de variación. La primera, que siempre está presente, se debe al error aleatorio en la medida, lo cual se analizó con detalle en el capítulo anterior: este tipo de error provoca diferentes resultados aunque las medidas se repitan en las mismas condiciones. La segunda fuente de variación posible se debe a lo que se conoce como **factor controlado** o **de efecto fijo**: para los ejemplos anteriores los factores de control son, respectivamente, las condiciones bajo las cuales se almacenó la solución, el método de análisis empleado y los operadores que realizaron la titulación. **El análisis de la varianza** (conocido como **ANOVA**, del inglés *Analysis of Variance*) es una técnica estadística muy poderosa que se utiliza para separar y estimar las diferentes causas de variación. En los ejemplos anteriores, se puede utilizar para separar la variación debida al error aleatorio de

cualquier otra variación provocada al cambiar el factor de control. De esta manera, probar si una alteración del factor de control ocasiona diferencias significativas entre los valores medios obtenidos.

También se puede emplear las técnicas ANOVA en situaciones donde hay más de una fuente de variación aleatoria. Por ejemplo, considere la pureza de un barril de cloruro sódico. Las muestras se toman al azar de diferentes partes del barril, y se realizan los mismos análisis sobre las muestras. Además del error aleatorio en la medición de la pureza, habrá también diferencias en la pureza de las muestras en las distintas partes del barril. Puesto que las muestras se eligen al azar, esta variación será aleatoria y esto se conoce como **factor de efecto aleatorio**. De nueva cuenta, en este caso puede utilizarse el ANOVA para separar las fuentes de variación.

Ambos tipos de análisis estadísticos, en donde hay *un* factor, ya sea controlado o aleatorio, además del error aleatorio de las medidas, se conoce como **ANOVA de un factor**. Los procedimientos aritméticos son similares en los casos de un factor de efecto fijo y uno de efecto aleatorio. En este capítulo se presentan ejemplos del primer caso y en el próximo capítulo del segundo, en donde se considera con más detalle el muestreo. Las situaciones más complejas en las que existen dos o más factores, posiblemente interactuando entre sí, se consideran en el capítulo 7 (Diseño experimental).

### 3.9 COMPARACIÓN DE VARIAS MEDIAS

En la tabla 3.1 se muestran los resultados obtenidos en una investigación acerca de la estabilidad de un reactivo fluorescente en diferentes condiciones de almacenamiento. Los valores dados son las señas de fluorescencia (en unidades arbitrarias) de soluciones diluidas de la misma concentración. Se tomaron tres medidas sobre cada muestra. En la tabla se muestra que los valores medios son diferentes para las cuatro muestras. Sin embargo, sabemos que, debido al error aleatorio, incluso si no ha cambiado el valor verdadero que queremos medir, la media muestral puede variar de una muestra a otra. Con las pruebas ANOVA se verifica si la diferencia entre las medias muestrales es demasiado grande para explicarse por medio de los errores aleatorios.

**Tabla 3.1** Señales de fluorescencia para soluciones en diferentes condiciones de almacenamiento

Condiciones	Medidas repetidas	Media
A recientemente preparada	102, 100, 101	101
B una hora en la oscuridad	101, 101, 104	102
C una hora con luz tenue	97, 95, 99	97
D una hora con luz brillante	90, 92, 94	92
	Media global	98

El problema se puede generalizar si se consideran *h* muestras, cada una con *n* elementos, como se expone en la tabla 3.2, en donde *x<sub>ij</sub>* es la *j*-ésima medida de la *i*-ésima muestra. Las medias de las muestras son  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_h$  y la media de todos los valores agrupados juntos es  $\bar{x}$ . La hipótesis nula adoptada es que todas las muestras se extraen de

**Tabla 3.2**
Generalización de la tabla 3.1

							Media
Muestra 1	$x_{11}$	$x_{12}$	. . . . .	$x_{1j}$	. . . . .	$x_{1n}$	$\bar{x}_1$
Muestra 2	$x_{21}$	$x_{22}$	. . . . .	$x_{2j}$	. . . . .	$x_{2n}$	$\bar{x}_2$
Muestra <i>i</i>	$x_{i1}$	$x_{i2}$	. . . . .	$x_{ij}$	. . . . .	$x_{in}$	$\bar{x}_i$
Muestra <i>h</i>	$x_{h1}$	$x_{h2}$	. . . . .	$x_{hj}$	. . . . .	$x_{hn}$	$\bar{x}_h$
							media global = $\bar{x}$

una población con media  $\mu$  y varianza  $\sigma_0^2$ . Con base en esta hipótesis se puede estimar  $\sigma_0^2$  de dos formas; una estudia la variación *dentro* de cada muestra, y la otra la variación *entre* las distintas muestras.

(i) *Variación dentro de la muestra*  
Para cada muestra se puede calcular una varianza utilizando la fórmula  $\sum (x_i - \bar{x})^2 / (n - 1)$ . Si usamos los valores de la tabla 3.1, tenemos que las varianzas de las muestras son:

Muestra A

$$= \frac{(102 - 101)^2 + (100 - 101)^2 + (101 - 101)^2}{3 - 1}$$

$$= 1$$

Muestra B

$$= \frac{(101 - 102)^2 + (101 - 102)^2 + (104 - 102)^2}{3 - 1}$$

$$= 3$$

Muestra C

$$= \frac{(97 - 97)^2 + (95 - 97)^2 + (99 - 97)^2}{3 - 1}$$

$$= 4$$

Muestra D

$$= \frac{(89 - 91)^2 + (91 - 91)^2 + (93 - 91)^2}{3 - 1}$$

$$= 4$$

Al promediar estos valores resulta:

estimación de  $\sigma_0^2$  dentro de la muestra

$$= (1 + 3 + 4 + 4) / 4 = 3$$

Esta estimación tiene 8 grados de libertad: la estimación de la varianza de cada muestra tiene 2 grados de libertad y hay 4 muestras. Destaca el hecho de que esta estimación no depende de las medias de las muestras; por ejemplo, si todas las medidas de la muestra A se incrementaran en cuatro unidades, no variaría la estimación de  $\sigma_0^2$ .

La fórmula general para la estimación de  $\sigma_0^2$  dentro de las muestras es:

$$\text{estimación dentro de las muestras de } \sigma_0^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / h(n-1) \quad (3.9)$$

El sumatorio sobre  $j$  y la división por  $(n-1)$  se refiere a la varianza de cada muestra; el sumatorio sobre  $i$  y la división por  $h$  promedia estas varianzas muestrales. La expresión en la ecuación 3.9 se conoce como un **cuadrado medio**, puesto que es la suma de términos cuadráticos dividida por el número de grados de libertad. Ya que en este caso el número de grados de libertad es 8 y el cuadrado medio es 3, la suma de los términos cuadráticos es  $3 \times 8 = 24$ .

## (ii) Variación entre muestras

Si todas las muestras se extraen de una población cuya varianza es  $\sigma_0^2$ , entonces sus medias proceden de una población con varianza  $\sigma_0^2/n$  (véase la distribución muestral de la media, en la sección 2.3). Así, si la hipótesis nula es verdadera, la varianza de las medias de las muestras proporciona una estimación de  $\sigma_0^2/n$ . De la tabla 3.1:

$$\begin{aligned} \text{varianza de la media muestral} &= \frac{(101 - 98)^2 + (102 - 98)^2 + (97 - 98)^2 + (92 - 98)^2}{4 - 1} \\ &= \frac{62}{3} \end{aligned}$$

de manera que:

$$\text{estimación entre muestras de } \sigma_0^2 = \frac{62}{3} \times 3 = 62$$

Esta estimación tiene 3 grados de libertad, ya que se ha calculado a partir de 4 medias muestrales. Note que esta estimación de  $\sigma_0^2$  no depende de la variabilidad *dentro* de cada muestra, ya que se calcula con las medias muestrales. Sin embargo, si por ejemplo la media de la muestra D cambiase entonces también cambiaría esta estimación de  $\sigma_0^2$ .

En general, tenemos:

$$\text{estimación de } \sigma_0^2 \text{ entre muestras} = n \sum_i (\bar{x}_i - \bar{x})^2 / (h-1) \quad (3.10)$$

que de nuevo es un "cuadrado medio" el cual supone dividir una suma de términos cuadráticos entre el número de grados de libertad. En este caso, el número de grados de libertad es 3 y el cuadrado medio es 62, por lo que la suma de los términos cuadráticos es  $3 \times 62 = 186$ .

Si resumimos los cálculos realizados hasta aquí:

cuadrado medio dentro de las muestras = 3 con 8 g.l.

cuadrado medio entre muestras = 62 con 3 g.l.

Si la hipótesis nula es correcta, estas dos estimaciones de  $\sigma_0^2$  no deberían de diferir en forma significativa. Si es incorrecta, la estimación de  $\sigma_0^2$  entre muestras será mayor que la estimación dentro de las muestras, debido a la variación entre muestras. Para probar si

la estimación entre muestras es significativamente más grande, se utiliza una prueba  $F$  de una cola (véase la Sec. 3.6):

$$F_{3,8} = 62/3 = 20.7$$

(Recuerde que se utiliza cada *cuadrado* medio, así que no es necesario elevar al cuadrado posteriormente.) De la tabla A.2 se obtiene que el valor crítico de  $F$  es 4.066 ( $P = 0.05$ ). Puesto que el valor calculado de  $F$  es mayor que éste, la hipótesis nula se rechaza: las medias muestrales *difieren* en forma significativa.

Se puede alcanzar un resultado significativo en un ANOVA de un factor por varias razones: por ejemplo, que una media difiera de las demás; que todas las medias difieran entre sí; que las medias estén en dos grupos distintos, etc. Una forma sencilla para explicar un resultado significativo es colocar las medias en orden creciente y comparar la diferencia entre los valores adyacentes con una cantidad llamada la **diferencia significativa menor**. Ésta está dada por  $s\sqrt{2/n} \times t_{h(n-1)}$ , en donde  $s$  es la estimación dentro de las muestras de  $\sigma_0$  y  $h(n-1)$  es el número de grados de libertad de esta estimación. Para el ejemplo anterior, las medias muestrales colocadas en orden creciente de tamaño son:

$$\bar{x}_D = 92 \quad \bar{x}_C = 97 \quad \bar{x}_A = 101 \quad \bar{x}_B = 102$$

y la menor diferencia significativa es  $\sqrt{3} \times \sqrt{(2/3)} \times 2.306$  ( $P = 0.05$ ), lo cual da 3.26. Al comparar este valor con las diferencias entre las medias, se sugiere que  $\bar{x}_D$  y  $\bar{x}_C$  difieren significativamente entre sí y también difieren de  $\bar{x}_A$  y  $\bar{x}_B$ , mientras que  $\bar{x}_A$  y  $\bar{x}_B$  no difieren en forma significativa entre sí, es decir, la exposición a la luz es lo que afecta la fluorescencia.

El método de la mínima diferencia significativa descrito antes no es del todo riguroso: se puede demostrar que conduce a demasiadas diferencias significativas. Sin embargo, es una prueba de ejecución simple cuando el ANOVA indica que *hay* diferencias significativas entre las medias. Otros métodos más rigurosos se describen en la bibliografía que aparece al final de este capítulo.

## 3.10 LA ARITMÉTICA DE LOS CÁLCULOS ANOVA

Al utilizar ANOVA para probar la diferencia entre varias medias, se estimó  $\sigma_0^2$  de dos formas distintas. Si la hipótesis nula fuese verdadera,  $\sigma_0^2$  también podría estimarse de una tercera forma, tratando los datos como si fueran una muestra grande. Esto implicaría sumar los cuadrados de las desviaciones a la media total:

$$\begin{aligned} \sum_i \sum_j (x_{ij} - \bar{x})^2 &= 4^2 + 2^2 + 3^2 + 3^2 + 3^2 + 6^2 + 1^2 + 3^2 + 1^2 + 8^2 + 6^2 + 4^2 \\ &= 210 \end{aligned}$$

y dividir por el número de grados de libertad,  $12 - 1 = 11$ .

Este método de estimar  $\sigma_0^2$  no se utiliza en el análisis debido a que la estimación depende de las dos variaciones: dentro y entre muestras. Sin embargo, existe una relación algebraica exacta entre esta variación total y las fuentes de variación que contribuyen a ella la cual, especialmente en los cálculos más complicados de ANOVA, conduce a una simplificación

**Tabla 3.3** Resumen de sumas de cuadrados y grados de libertad

Fuente de variación	Suma de cuadrados	Grados de libertad
Entre muestras	$n \sum_i (\bar{x}_i - \bar{x})^2 = 186$	$h - 1 = 3$
Dentro de las muestras	$\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 = 24$	$h(n - 1) = 8$
Total	$\sum_i \sum_j (x_{ij} - \bar{x})^2 = 210$	$hn - 1 = 11$

del trabajo aritmético. En la tabla 3.3 en donde se resumen las sumas de cuadrados y los grados de libertad, se muestra la relación entre las fuentes de variación. Se observará que los valores de la variación total, dados en la última fila de la tabla, son la suma de los valores de las dos primeras filas, tanto para la suma de cuadrados como para los grados de libertad. Esta propiedad aditiva se mantiene para todos los cálculos de ANOVA descritos en este libro.

De la misma manera que el cálculo de la varianza, hay fórmulas que simplifican el cálculo de la suma de cuadrados. Estas fórmulas se resumen en la tabla 3.4 en la cual se utiliza la notación anterior y también se introducen los símbolos:

número total de medidas =  $N = nh$

suma de las medidas en la  $i$ -ésima muestra =  $T_i$

suma de todas las medidas, gran total =  $T$

**Tabla 3.4** Fórmula para cálculos de ANOVA de un factor

Fuente de variación	Suma de cuadrados	Grados de libertad
Entre muestras	$\sum_i T_i^2/n - T^2/N$	$h - 1$
Dentro de las muestras	por diferencia	por diferencia
Total	$\sum_i \sum_j x_{ij}^2 - T^2/N$	$N - 1$

Se puede entender el funcionamiento de las fórmulas en la tabla 3.4 si se repiten los cálculos de ANOVA para los datos de la tabla 3.1. El cálculo de los cuadrados de la media se expone más adelante. A todos los valores de la tabla 3.1 se les ha restado 100, lo que simplifica considerablemente la aritmética. Note que esto no afecta la estimación de la varianza entre muestras ni dentro de las muestras, ya que se ha restado la misma cantidad de *cada* valor.

A	2	0	1	$T_i$	$T_i^2$
B	1	1	4	3	9
C	-3	-5	-1	6	36
D	-10	-8	-6	-9	81
				-24	576
				$T = -24$	$\sum T_i^2 = 702$

$$n = 3 \quad h = 4 \quad N = 12 \quad \sum_i \sum_j x_{ij}^2 = 258$$

Fuente de variación	Suma de los cuadrados	Grados de libertad	Cuadrado medio
Entre muestras	$702 - (-24)^2/12 = 186$	3	$186/3 = 62$
Dentro de las muestras	por diferencia = 24	8	$24/8 = 3$
Total	$258 - (-24)^2/12 = 210$	11	

### 3.11 LA PRUEBA CHI CUADRADA

Las pruebas de significación descritos hasta ahora en este capítulo han sido, en general, pruebas acerca de si la media de varias observaciones difiere significativamente del valor propuesto por la hipótesis nula. Los datos utilizados han tomado la forma de observaciones que, aparte de cualquier redondeo, se han medido en una escala continua. En contraposición en esta sección se tratan las *frecuencias* es decir, el número de veces que ocurre un suceso dado. Por ejemplo, en la tabla 2.3 se proporcionan las frecuencias de los diferentes valores obtenidos de la concentración de ion nitrato cuando se tomaron 50 medidas en una muestra. Como se analizó en el capítulo 2, por lo regular se supone que estas medidas se extraen de una población que se distribuye normalmente: la prueba chi cuadrada se puede utilizar para probar si las frecuencias observadas difieren en forma significativa de las que cabría esperar con esta hipótesis nula. Puesto que el cálculo requerido en este caso es relativamente complicado, no será descrito aquí. (Al final del capítulo se proporciona un ejemplo ilustrativo desarrollado con detalle.) El principio de la prueba chi cuadrada se comprende mejor en el siguiente ejemplo:

*Ejemplo.* A continuación se muestra el número de roturas en el material de vidrio informadas por 4 trabajadores de un laboratorio a lo largo de cierto periodo. ¿Hay alguna prueba de que los trabajadores difieran en su habilidad?

Número de roturas: 24, 17, 11, 9.

La hipótesis nula adoptada es que no hay diferencias en su habilidad. Si los trabajadores utilizan el laboratorio el mismo tiempo, se esperaría, a partir de la hipótesis nula, el mismo número de roturas por cada trabajador. Puesto que el número total de roturas es 61, el número de roturas esperada por trabajadores es  $61/4 = 15.25$ . Por supuesto, en la práctica no es posible que haya un número no entero de roturas: éste corresponde a un concepto matemático. La distribución "igual" más próxima y manejable desde un punto de vista práctico es 15, 15, 15, 16, en algún orden determinado. Se trata de responder la pregunta de si la diferencia entre las frecuencias observadas y esperadas es tan grande que deba rechazarse la hipótesis nula. Si consideramos una serie de lanzamientos de un dado se puede entender mejor por qué existen *ciertas* diferencias entre los dos conjuntos de frecuencias: sería muy sorprendente, por ejemplo, si en 30 lanzamientos aparecieran exactamente con la misma frecuencia los números 1, 2, 3, etc. El cálculo de chi cuadrada,  $\chi^2$ , que es la cantidad utilizada para probar la existencia de diferencias significativas, se muestra a continuación.

Frecuencia observada,	Frecuencia esperada, $E$	$O - E$	$(O - E)^2/E$
24	15.25	8.75	5.020
17	15.25	1.75	0.201
11	15.25	-4.25	1.184
9	15.25	-6.25	2.561
		0.00	$\chi^2 = 8.966$

Note que el total de la columna  $O - E$  es siempre cero, con lo cual se dispone de una forma útil para verificar los cálculos.

Si  $\chi^2$  supera un cierto valor crítico, se rechaza la hipótesis. El valor crítico depende, como en otras pruebas de significación, del nivel de significación de la prueba y del número de grados de libertad. En un ejemplo de este tipo, el número de grados de libertad es uno menos que el número de clases utilizado, es decir,  $4 - 1 = 3$  en este caso. Los valores críticos de  $\chi^2$  para  $P = 0.05$  aparecen en la tabla A.5. Para 3 grados de libertad, el valor crítico es 7.81. Puesto que el valor calculado de  $\chi^2$  es mayor que éste se rechaza la hipótesis nula a un nivel de significación del 5%: se han encontrado pruebas de que los trabajadores *difieren* significativamente en su habilidad.

En este cálculo de  $\chi^2$  se obtiene un resultado significativo debido al alto número de roturas realizadas por el primer trabajador. Para estudiar esto más ampliamente se pueden realizar pruebas chi cuadrada adicionales. Una de ellas prueba si el segundo, tercero y cuarto trabajadores difieren significativamente entre sí: en este caso, cada frecuencia esperada es  $(17 + 11 + 9)/3$ . (Note el hecho de que la prueba  $t$  no puede aplicarse aquí, ya que estamos manejando frecuencias y no variables continuas.) Otras pruebas que permiten analizar si el primer trabajador difiere del resto, es tomar a los tres restantes como un grupo. En este caso hay dos clases: las roturas hechas por el primer trabajador con una frecuencia esperada de 15.25 y el total de roturas correspondientes a los otros tres: con una frecuencia esperada de  $15.25 \times 3 = 45.75$ . En estos casos, donde hay sólo dos grupos y, por lo tanto, un grado de libertad, se debe aplicar un ajuste conocido como **corrección de Yates**; esto implica sustituir  $O - E$  por  $|O - E| - 0.5$ , por ejemplo, 4.5 se convierte en 4. Estas pruebas adicionales se muestran en un ejercicio al final de este capítulo.

En general, la prueba chi cuadrada debería utilizarse solamente si el número total de observaciones es 50 o más y las frecuencias individuales esperadas no son menores que 5. Esto no es una regla rígida: al final del capítulo aparece una referencia en la que se profundiza en este punto, y se describen también otras aplicaciones de la prueba chi cuadrada.

### 3.12 PRUEBA DE LA NORMALIDAD DE UNA DISTRIBUCIÓN

Como se ha señalado en este capítulo, muchas pruebas estadísticas suponen que los datos utilizados proceden de una población normal. En la sección anterior se mencionó un método para probar esta hipótesis, utilizando la prueba chi cuadrada. Por desgracia, este método sólo se puede utilizar si hay 50 o más datos. En el trabajo experimental lo normal es tener pocos datos. Una forma visual simple de comprobar si un grupo de datos procede de una población normal es representar una **curva de frecuencias acumuladas**

en un papel gráfico especial denominado **papel de probabilidad normal**. Este método se explica mejor con un ejemplo.

*Ejemplo.* Utilice papel de probabilidad normal para probar si los valores siguientes proceden de una población normal:

109, 89, 99, 99, 107, 111, 86, 74, 115, 107, 134, 113, 110, 88, 104.

En la tabla 3.5 se muestran los datos acomodados en orden creciente. En la segunda columna aparecen las frecuencias acumuladas en cada medida, es decir, el número de

Tabla 3.5

Medida	Frecuencia acumulada	Frecuencia acumulada en %
74	1	6.3
86	2	12.5
88	3	18.8
89	4	25.0
99	6	37.5
104	7	43.8
107	9	56.3
109	10	62.5
110	11	68.8
111	12	75.0
113	13	81.3
115	14	87.5
134	15	93.8

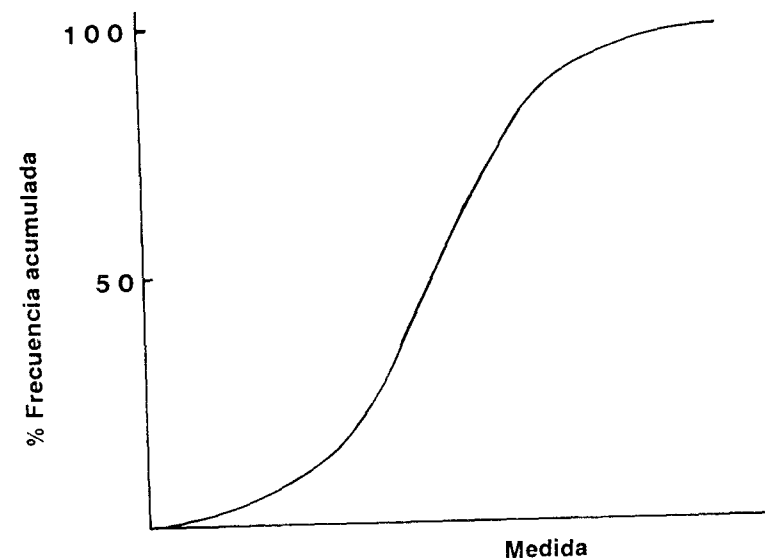


Figura 3.1 La curva de frecuencias acumuladas de una distribución normal.

medidas menores o iguales que dicha medida. En la tercera columna se muestra la frecuencia acumulada en porcentaje; éste se ha calculado utilizando la fórmula:

$$\% \text{ frecuencia acumulada} = 100 \times \text{frecuencia acumulada} / (n + 1)$$

Donde  $n$  es el número total de medidas (existen razones matemáticas para dividir por  $n + 1$  en lugar de  $n$ ). Si los datos proceden de una distribución normal, la gráfica que presenta la frecuencia acumulada frente a las medidas tendrá forma de curva en S, como se muestra en la figura 3.1.

El papel de probabilidad normal tiene una escala no lineal en el eje del porcentaje de frecuencia acumulada, lo que convierte la curva en forma de S en una línea recta. Los datos de la tabla anterior, representados en dicho papel, aparecen en la figura 3.2: los puntos se sitúan aproximadamente sobre una línea recta, lo cual confirma la hipótesis de que los datos proceden de una distribución normal.

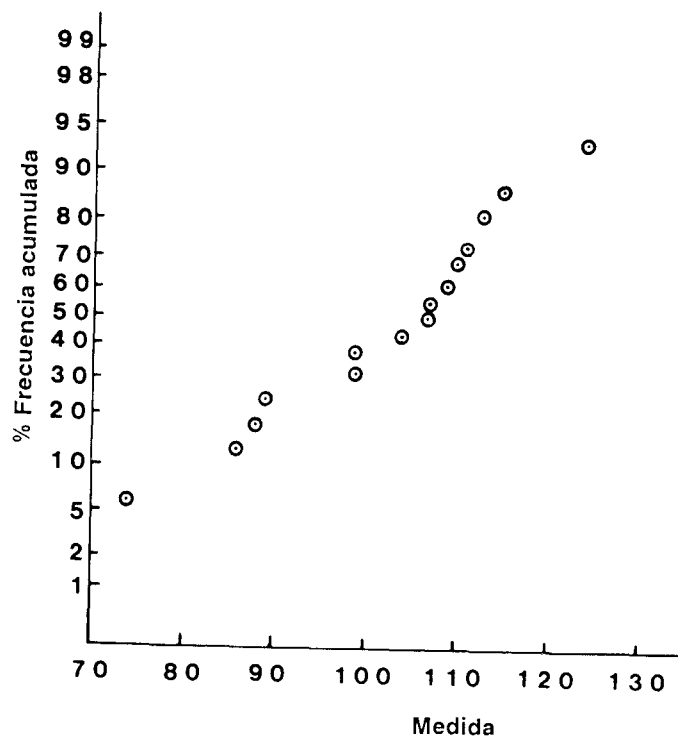


Figura 3.2 Representación de la probabilidad normal para el ejemplo de la sección 3.12.

Un método para probar la normalidad es utilizar una cantidad que mida cuánto se aproximan a una recta los puntos representados en el papel de probabilidad normal. El cálculo de esta cantidad, el coeficiente de correlación,  $r$ , se describe en un capítulo posterior (Sec. 5.3). Al final de este capítulo aparece una referencia bibliográfica para el uso de  $r$  al

probar la normalidad, en ella se proporciona también una perspectiva de las diferentes pruebas de normalidad. En la sección 6.11 se describe otro método, el método de Kolmogorov-Smirnov que, entre otras aplicaciones, se puede emplear para probar la normalidad. En esa sección aparece un ejemplo desarrollado.

### 3.13 CONCLUSIONES A PARTIR DE LAS PRUEBAS DE SIGNIFICACIÓN

Hasta ahora, en este capítulo nos hemos dedicado a los diferentes tipos de prueba de significación. En esta sección profundizaremos más en lo referente a las conclusiones que se pueden extraer de una prueba de significación. Como se explicó en la sección 3.2, una prueba de significación, por ejemplo, al nivel  $P = 0.05$ , supone un riesgo del 5% de que se rechazará una hipótesis nula *incluso aunque sea verdadera*. A este tipo de error se le conoce como **error de tipo 1**: el riesgo de este error se puede reducir al alterar el nivel de significación de la prueba a  $P = 0.01$  o incluso a  $P = 0.001$ . Sin embargo, éste no es el único tipo de error posible: también se puede *mantener* una hipótesis nula *aun cuando sea falsa*. Este es el denominado **error de tipo 2**. Para calcular la probabilidad de este tipo de error es necesario postular una alternativa a la hipótesis nula, conocida como **hipótesis alternativa**.

Considere un producto químico del cual se piensa que contiene el 3% de fósforo por pesada. Al parecer, esta proporción se ha incrementado y, para probar esto, se ha analizado la composición empleando un método estándar con una desviación estándar conocida del 0.03%. Suponga que se toman 4 medidas y se realiza una prueba de significación al nivel  $P = 0.05$ . Es necesario que dicha prueba sea de una cola, al estar interesados sólo en un *incremento*. La hipótesis nula es:

hipótesis nula, = 3.0%

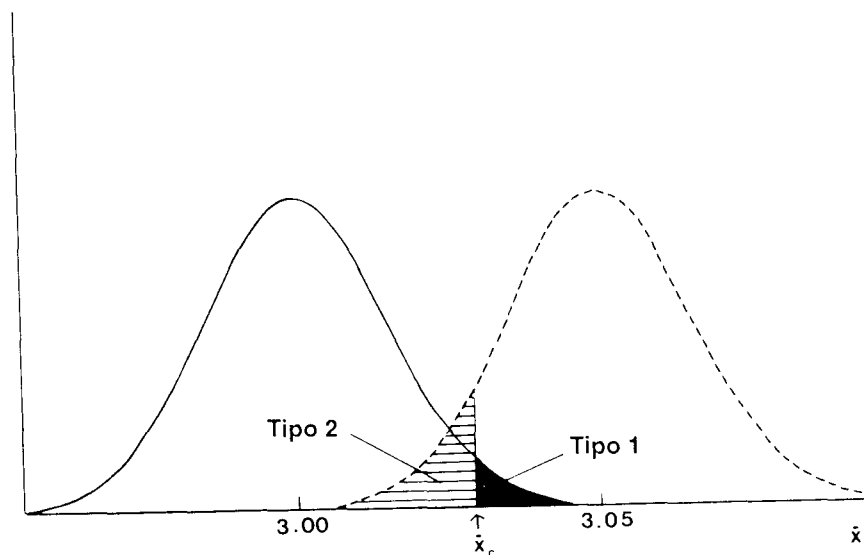


Figura 3.3 Errores de tipos 1 y 2.

La línea continua de la figura 3.3 indica la distribución muestral de la media si la hipótesis nula fuese verdadera. Esta distribución muestral tiene una media 3.0 y una desviación estándar (es decir, error estándar de la media)  $\sigma/\sqrt{n} = 0.03/\sqrt{4} \%$ . Si la media muestral cae por encima del valor crítico indicado,  $\bar{x}_c$ , se rechaza la hipótesis nula. Por lo tanto, la región sombreada, con área 0.05, representa la probabilidad de un error de tipo 1.

Supongamos que tenemos la hipótesis alternativa:

hipótesis alternativa,  $\mu = 3.05\%$

La línea punteada de la figura 3.3 indica la distribución muestral de la media si la hipótesis alternativa fuese verdadera. *incluso si éste fuese el caso*, se aceptará la hipótesis nula si la media muestral cae por debajo de  $\bar{x}_c$ . La probabilidad de este error de tipo 2 se representa con el área rayada. El diagrama pone de manifiesto la interdependencia de los tipos de error. Por ejemplo, si se cambia el nivel de significación a  $P = 0.01$  para reducir un riesgo de un error de tipo 1,  $\bar{x}_c$  aumentará y el riesgo de un error de tipo 2 se incrementará también. De manera recíproca, una disminución en el riesgo de un error de tipo 2 sólo puede lograrse por un incremento en la probabilidad de un error de tipo 1. La única manera por la que pueden reducirse ambos errores (para una hipótesis alternativa dada) es aumentar el tamaño de la muestra. Por ejemplo, el efecto de incremento  $n$  hasta 9, se ilustra en la figura 3.4: la disminución resultante en el error estándar de la media produce una disminución de ambos tipos de error, para un valor de  $\bar{x}_c$ .

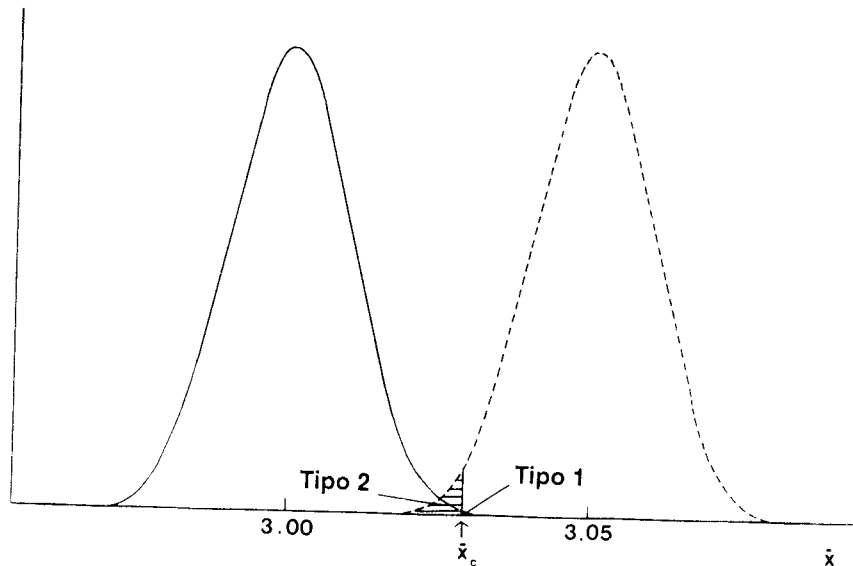


Figura 3.4 Errores de tipo 1 y 2 al aumentar el tamaño de la muestra.

La probabilidad de que se rechace una hipótesis nula falsa se denomina la **potencia** de una prueba; esto es  $(1 - \text{probabilidad de un error de tipo 2})$ . En el ejemplo expuesto es una función de la media especificada en la hipótesis alternativa, y depende también del tamaño de la muestra, el nivel de significación de la prueba, y si ésta es de una o dos colas. En

circunstancias en donde se dispone de dos o más pruebas para probar la misma hipótesis, puede ser útil comparar la potencias de las pruebas con el fin de decidir cuál es la más adecuada.

Los errores de tipo 1 y 2 son también relevantes cuando se aplican secuencialmente pruebas de significación. Un ejemplo de esta situación es la aplicación de la prueba  $t$  a la diferencia entre dos medias, después de utilizar en primer lugar la prueba  $F$  para decidir si se pueden juntar o no las varianzas de las muestras (véanse las Secs. 3.3 y 3.6). Los dos errores de tipo 1 y 2 pueden surgir de la prueba  $F$  inicial; si ocurre alguno de ellos querrá decir que los niveles de significación establecidos para la subsiguiente prueba  $t$  son incorrectos, debido a la forma incorrecta en que ha sido aplicada la prueba  $t$ .

Este ejemplo resalta la conclusión general de que las pruebas de significación no dan respuestas indiscutibles: más bien ayudan a la interpretación de los datos experimentales al dar las probabilidades de que ciertas conclusiones sean válidas.

## BIBLIOGRAFÍA

- Barnett, V. y Lewis, T., *Outliers in Statistical Data*, Nueva York, Wiley, 1978. Un tratamiento muy completo de la filosofía del rechazo de valores anómalos y las pruebas empleadas.
- Box, G. E. P., Hunter, W. G. y Hunter, J. S., *Statistics for Experimentalists*, Nueva York, Wiley, 1978. Proporciona más detalles sobre los contrastes entre medias en el ANOVA.
- Campbell, R. C., *Statistics for Biologists*, Cambridge, Cambridge University Press, 1974. Proporciona pruebas de la homogeneidad de varianzas y de normalidad.
- Davies, O. L. y Goldsmith, P. L., *Statistical Methods in Research and Production*, Londres, Longmans, 1972. Proporciona más detalles acerca de los errores de Tipos 1 y 2 y otras aplicaciones de la prueba chi cuadrada.
- Filliben, J. J., *Technometrics*, 1975, **17**, 111. Describe el uso de  $r$  al probar la normalidad, y estudia otras pruebas de normalidad.
- Kowalski, B. R. (Ed.), *Chemometrics: Theory and Application*, American Chemical Society, 1977, Washington. El capítulo 11 describe pruebas de normalidad y analiza el efecto de la no normalidad en las pruebas paramétricas.
- Miller, J. C., *Statistics for Advanced Level*, Cambridge, Cambridge University Press, 1983. Proporciona un ejemplo de la prueba chi cuadrada para normalidad.
- Sokal, R. R. y Rohlf, F. J., *Biometry*, Freeman, 1969. Da información detallada de las pruebas de la homogeneidad de varianzas.

## EJERCICIOS

- Utilizando una gráfica de probabilidad normal, pruebe si el siguiente conjunto de datos podría haber sido obtenido de una población normal:  
  
11.68, 11.12, 8.92, 8.82, 10.31, 11.88, 9.84, 11.69, 9.53, 10.30, 9.17, 10.04, 10.65, 10.91, 10.32, 8.71, 9.83, 8.90, 10.40
- Los datos expuestos a continuación están tomados del ejemplo de la sección 3.3 y se refieren a la concentración de tiol en sangre. Verifique si el dato 2.07 no es un valor anómalo.  
  
1.84 1.92 1.94 1.85 1.91 2.07



3. Los siguientes datos proporcionan la recuperación de bromuro adicionado a muestras con contenido vegetal, medido mediante un método de cromatografía gas-líquido. La cantidad de bromuro potásico añadido a cada vegetal fue la misma.

Tomate 777 790 759 790 770 758 764  $\mu\text{g/g}$   
 Pepino 782 773 778 765 789 797 782  $\mu\text{g/g}$

(Roughan, J. A., Roughan, P. A. y Wilkins, J. P. G., *Analyst*, 1983, **108**, 742.)

- (a) Pruebe si la recuperación en los vegetales tiene varianzas que difieran significativamente.  
 (b) Pruebe si las tasas de recuperación medias difieren significativamente.
4. Los siguientes resultados muestran el porcentaje del agua intersticial total recuperada al centrifugar muestras de piedras areniscas tomadas de diferentes profundidades.

Profundidad de la muestra (m)	Agua recuperada, %							
7	33.3	33.3	35.7	38.1	31.0	33.3		
8	43.6	45.2	47.7	45.4	43.8	46.5		
16	73.2	68.7	73.6	70.9	72.5	74.5		
23	72.5	70.4	65.2	66.7	77.6	69.8		

(Wheatstone, K. G. y Gelsthorpe, D., *Analyst*, 1982, **107**, 731.)

Demuestre que el porcentaje de agua recuperada difiere significativamente de diferentes profundidades. Utilice el método de la menor diferencia significativa descrito en la sección 3.9 para encontrar las causas de este resultado significativo.

5. Al leer en una bureta 0.01 ml, un analista tiene que estimar la cifra final. La siguiente tabla de frecuencias proporciona las últimas cifras de 50 lecturas. Realice una prueba de significación adecuada para determinar si se prefieren algunos dígitos a otros.

Dígito	0	1	2	3	4	5	6	7	8	9
Frecuencia	1	6	4	5	3	11	2	8	3	7

6. Para evaluar un método espectrofotométrico con el fin de determinar titanio, se aplicó el método a muestras de aleaciones conteniendo diferentes cantidades certificadas de titanio. Los resultados (% Ti) se muestran a continuación.

Muestra	Valor certificado	Media	Desviación estándar
1	0.496	0.482	0.0257
2	0.995	1.009	0.0248
3	1.493	1.505	0.0287
4	1.990	2.002	0.0212

Para cada aleación se realizaron 8 determinaciones repetidas.

(Qiu Xing-chu y Zhu Ying-quen, *Analyst*, 1983, **108**, 641.)

Para cada aleación pruebe si el valor medio difiere significativamente del valor certificado.

7. La siguiente tabla muestra otros resultados del artículo citado en la sección 3.3, concerniente a la extracción y determinación de estaño en comestibles (*Analyst*, 1983, **108**, 109). Los resultados proporcionan los niveles de estaño recuperado del mismo producto después de ebullición a diferentes tiempos en un recipiente abierto.

Tiempo de ebullición (min)	Estaño encontrado (mg/kg)
30	57, 57, 55, 56, 56, 55, 56, 55
75	51, 60, 48, 32, 46, 58, 56, 51

Pruebe si

- (a) la variabilidad de los resultados es más grande para tiempos de ebullición más largos.  
 (b) las medias difieren significativamente.

8. Los datos dados en el ejemplo de la sección 3.12, para el número de roturas realizadas por cuatro trabajadores diferentes, se reproducen a continuación:

24, 17, 11, 9

Pruebe si

- (a) el número de roturas del primer trabajador difieren significativamente de las de los otros tres trabajadores;  
 (b) el segundo, tercero y cuarto trabajadores difieren significativamente en la falta de cuidado de cada uno de los otros.

9. Los siguientes resultados se obtuvieron al comprar un método nuevo y el método oficial para la determinación de fenilbutazona. (Las cifras se refieren a los porcentajes de recuperación.)

	Método nuevo	Método oficial
Media	99.35	99.53
Varianza	0.185	0.152
Tamaño de muestra	3	3

(Amer, M. M., Taha, A. H., El-Zeany, B. A. y El-Sawy, O. A., *Analyst*, 1982, **107**, 908.)

Pruebe si los resultados medios obtenidos por los dos métodos difieren significativamente.

10. La siguiente tabla proporciona la cantidad (en mg/ml) de clorhidrato de efedrina encontrada en preparaciones farmacéuticas de Ephedrine Elixir B. P., por dos métodos diferentes: espectroscopía ultravioleta de derivadas y un método de ensayo oficial. (La cantidad nominal en cada muestra fue de 3 mg/ml.)

Muestra número	Método de derivadas	Método oficial
1	2.964	2.913
2	3.030	3.000
3	2.994	3.024

(Davidson, A. G. y Elsheikh, H., *Analyst*, 1982, **107**, 879.)

Pruebe si los resultados obtenidos por ambos métodos difieren significativamente.

11. Las siguientes cifras se refieren a la concentración de albúmina, en g/l., en el suero sanguíneo de 16 adultos sanos:

37, 39, 37, 42, 39, 45, 42, 39, 44, 40, 39, 45, 47, 47, 43, 41

(Foote, J. W. y Delves, H. T., *Analyst*, 1983, **108**, 492.)

Las primeras 8 cifras corresponden a hombres y las segundas 8 cifras a mujeres. Pruebe si la concentración media para hombres y mujeres difiere significativamente.

12. Se comparó un nuevo método espectroscópico de absorción atómica de llama para determinar antimonio en la atmósfera con el método colorimétrico recomendado. Para muestras de atmósfera urbana se obtuvieron los siguientes resultados:

Muestra número	Antimonio encontrado (mg/m <sup>3</sup> )	
	Método nuevo	Método estándar
1	22.2	25.0
2	19.2	19.5
3	15.7	16.6
4	20.4	21.3
5	19.6	20.7
6	15.7	16.8

(Castillo, J. R., Lanaja, J., Martínez, M. C. y Aznárez, J., *Analyst*, 1982, **107**, 1448.)

¿Difieren significativamente los resultados obtenidos por los dos métodos?

## Control de calidad y muestreo

### 4.1 INTRODUCCIÓN

Las pruebas estadísticas descritas en el capítulo anterior se aplicaron a situaciones mucho más sencillas que las que se presentan en realidad en los laboratorios analíticos. Entonces asumíamos que no se presentaba ninguna dificultad o error al obtener las muestras químicas que posteriormente serían suministradas para el análisis. En la práctica, el muestreo causa problemas en bastantes análisis directos. El análisis de boro en muestras de plantas (véase la Sec. 3.3) podría complicarse en la práctica si los niveles de boro varían en diferentes partes de una única planta, o de una planta a otra. También deben tratarse otros dos problemas: las variaciones existentes cuando se aplica el mismo método a una muestra o muestras similares en *diferentes laboratorios*, y la aplicación de la estadística a mediciones repetidas sobre muestras que tienen características que pueden *variar con el tiempo*, tales como los sucesivos artículos de una línea de producción. En estos estudios se aplican métodos estadísticos convencionales (p. ej. pruebas para valores anómalos, pruebas ANOVA, etc.) a situaciones que son muy importantes en el desarrollo y la aplicación de métodos analíticos.

### 4.2 MUESTREO

Este tema es de fundamental importancia porque si no se toma en cuenta la etapa de muestreo de un análisis con la consideración que merece, puede ser que los métodos estadísticos descritos en este libro no sean válidos al aplicarse a los resultados. Un analista tiene que tratar con muestras ya que en muchos casos es impracticable o imposible analizar todo el objeto en estudio. Por ejemplo, es impracticable analizar un tanque lleno de leche para examinar el contenido en grasa, y es imposible analizar toda el agua de un río para estudiar un determinado contaminante. Además, muchas técnicas analíticas son destructivas y no pueden aplicarse a la totalidad de un objeto de valor.

Para ilustrar algunos aspectos del muestreo consideremos la siguiente situación: disponemos de un gran lote de pastillas y deseamos obtener una estimación del peso medio

de una pastilla. En vez de pesar todas las pastillas, tomamos unas pocas (supongamos diez) y pesamos cada una de ellas. En este ejemplo el lote de tabletas forma la población y las diez pastillas pesadas constituyen una muestra de esta población (véase la Sec. 2.2). Si la muestra se va a utilizar para deducir las propiedades de la población, entonces debe ser lo que estadísticamente se denomina como una **muestra aleatoria**. Esta muestra es tomada de tal modo que todos los miembros de la población tienen la misma posibilidad de ser incluidos. Sólo entonces serán válidas las ecuaciones como la (2.9), que proporciona los límites de confianza de la media. Debemos apreciar que el término “aleatorio” tiene, en sentido estadístico, un significado diferente de “casual”. Aunque en la práctica un analista puede diseminar las pastillas sobre una mesa e intentar tomar una muestra de diez de una manera casual, dicho método podría ocultar un sesgo inconsciente. El mejor modo de obtener una muestra aleatoria es mediante el uso de una tabla de números aleatorios. A cada miembro de la población se le asigna un número de tal manera que todos los números tengan la misma cantidad de dígitos, por ejemplo, 001, 002, 003, etc. Entonces, los números aleatorios se leen de una tabla de números aleatorios (véase la tabla A.6), partiendo de un punto arbitrario que proporcione, por ejemplo, 964, 173, etc., y los números correspondientes de la población constituirán la muestra. Un procedimiento alternativo (y mucho más simple), que se utiliza algunas veces, es el de seleccionar los elementos de una población a intervalos regulares, por ejemplo, tomar una de cada cien pastillas de una línea de producción. Este procedimiento no es completamente satisfactorio, ya que podría existir una periodicidad coincidente en el peso de las pastillas. La importancia de la aleatoriedad de la muestra es evidente. De nuevo, si las últimas pastillas tomadas hubieran presentado una disminución gradual del peso durante la producción del lote, entonces esta muestra proporcionaría un valor completamente erróneo del peso medio del lote completo.

En el ejemplo anterior, la población está constituida por elementos discretos obvios, los cuales nominalmente son los mismos, o sea, las pastillas. El muestreo de materiales como rocas, polvo, gases y líquidos, para los que esto no se cumple, se denomina **muestreo a granel**. La falta de unidades identificables conduce a cierta confusión en la nomenclatura, entonces ¿debería considerarse “muestra” en el sentido coloquial de “una pequeña parte del conjunto”, de manera que *una* muestra correspondiese a *una* pastilla del ejemplo anterior, o debería considerar el significado estadístico de varias pequeñas partes tomadas de diferentes lugares del grueso de manera que *una* muestra correspondiese a *varias* tabletas? Para evitar esta confusión en el desarrollo posterior, a una pequeña parte tomada del grueso se le denominará **incremento de muestra** y a un grupo de tales incrementos se le llamará **muestra bruta**.

Si un material a granel fuera perfectamente homogéneo, sólo sería necesario un incremento de muestra para determinar las propiedades de la totalidad. En la práctica los materiales a granel no son homogéneos por diversas razones; por ejemplo, las aleaciones y los sedimentos están formados por partículas microscópicas con diferente composición y puede ser que no estén uniformemente distribuidas en el conjunto, o puede ser que los fluidos no sean homogéneos en una escala molecular, debido a los gradientes de concentración. Tal falta de homogeneidad se detecta sólo tomando incrementos de muestra de diferentes partes del conjunto. Para obtener una muestra bruta aleatoria deberían tomarse dichos incrementos en puntos del conjunto elegidos al azar. Esto se puede hacer si se considera el grueso o el conjunto como una colección de celdillas y si se seleccionan las celdillas que van a ser muestreadas, utilizando números aleatorios como se describió anteriormente.

De la muestra aleatoria pueden calcularse la media,  $\bar{x}$ , y la varianza  $s^2$ . Existen dos contribuciones al valor de  $s^2$ : la **varianza muestral**,  $\sigma_1^2$ , debida a las diferencias entre los elementos de la muestra, por ejemplo, pastillas que tienen distintos pesos, y la **varianza en la medición**,  $\sigma_0^2$ , por ejemplo, errores aleatorios al pesar cada pastilla. En la sección siguiente se describe cómo pueden separarse y estimarse estas dos contribuciones utilizando el ANOVA.

### 4.3 ESTIMACIÓN Y SEPARACIÓN DE VARIANZAS UTILIZANDO ANOVA

En la sección 3.8 se describió el uso de ANOVA de un factor para probar las diferencias entre medias cuando existe una posible variación debido a un factor de efecto fijo. En esta sección consideraremos la situación en donde existe un factor de efecto aleatorio, o sea, la variación en el muestreo; por lo tanto, el ANOVA de un factor no se utiliza para probar si varias medidas muestrales difieren significativamente, sino para separar y estimar las diferentes fuentes de variación. En la tabla 4.1 se muestran los resultados de la prueba de pureza de un barril de cloruro sódico. Se tomaron cinco incrementos de muestra, de A a E, de diferentes partes del barril elegidos al azar, y se realizaron cuatro análisis repetidos sobre cada muestra. Como se explicó anteriormente, hay dos fuentes de variación posibles: aquella debida al error aleatorio en la medición de la pureza, dada por la varianza en la medición,  $\sigma_0^2$ , y la que se debe a las variaciones reales en la pureza de cloruro sódico en diferentes puntos, dada por la varianza muestral,  $\sigma_1^2$ . Puesto que el cuadrado medio dentro de las muestras no depende de la media muestral (véase la Sec. 3.9) puede usarse para estimar  $\sigma_0^2$ . El cuadrado medio entre muestras *no se puede* utilizar para estimar  $\sigma_1^2$  en forma directa, ya que la variación entre medias muestrales se debe al error aleatorio en la medición y a las posibles variaciones en la pureza. Se puede demostrar que el cuadrado medio entre las muestras proporciona una estimación de  $\sigma_0^2 + n\sigma_1^2$  (donde, como en la Sec. 3.9,  $n$  es el número de mediciones repetidas). Sin embargo, antes de realizar una estimación de  $\sigma_1^2$  debería hacerse una prueba para ver si difiere significativamente de 0, comparando los cuadrados medios entre y dentro de las muestras: si no difieren de manera significativa entonces  $\sigma_1^2 = 0$  y ambos cuadrados medios estiman  $\sigma_0^2$ .

Tabla 4.1 Prueba de pureza de cloruro sódico

Muestra	Pureza (%)	Media
A	98.8, 98.7 98.9, 98.8	98.8
B	99.3, 98.7, 98.8, 99.2	99.0
C	98.3, 98.5, 98.8, 98.8	98.6
D	98.0, 97.7, 97.4, 97.3	97.6
E	99.3, 99.4, 99.9, 99.4	99.5

A continuación se desarrolla el cálculo de los cuadrados medios utilizando las fórmulas de la tabla 3.4. A todos los valores de la tabla 4.1 se les resta 98.5, para simplificar la aritmética.

					$T_i$	$T_i^2$
Muestra A	0.3	0.2	0.4	0.3	1.2	1.44
Muestra B	0.8	0.2	0.3	0.7	2.0	4.00
Muestra C	-0.2	0.0	0.3	0.3	0.4	0.16
Muestra D	-0.5	-0.8	-1.1	-1.2	-3.6	12.96
Muestra E	0.8	0.9	1.4	0.9	4.0	16.00
				$T = 4.0$	$\sum T_i^2 = 34.56$	

$$n = 4 \quad h = 5 \quad N = 20 \quad \sum \sum x_{ij}^2 = 9.62$$

Fuente de variación	Suma de cuadrados	Por diferencia	Grados de libertad
Entre las muestras	$34.56/4 - 4.0^2/20 = 7.84$	4	$7.84/4 = 1.96$
Dentro de las muestras	Cuadrados medios = 0.98	15	$0.98/15 = 0.0653$
Total	$9.62 - 4.0^2/20 = 8.82$	19	

Ya que el cuadrado medio entre las muestras es mayor que el cuadrado medio dentro de las muestras,  $\sigma_1^2$  puede diferir significativamente de 0. Utilizando la prueba  $F$  para comparar los dos cuadrados medios tenemos:

$$F_{4,15} = 1.96/0.0653 = 30$$

De la tabla A.2 se desprende que el valor crítico de  $F$  es 3.056 ( $P = 0.05$ ). El valor calculado de  $F$  es mayor, por lo tanto  $\sigma_1^2$  difiere significativamente de 0. El cuadrado medio entre las muestras da el valor de 0.0653 como una estimación de  $\sigma_0^2$ , puesto que el cuadrado medio entre muestras se estima como  $\sigma_0^2 + n\sigma_1^2$  tenemos:

$$\begin{aligned} \text{estimación de } \sigma_1^2 &= (\text{cuadrado medio entre muestras} - \text{cuadrado medio} \\ &\quad \text{dentro de las muestras})/n \\ &= (1.96 - 0.0653)/4 \\ &= 0.47 \end{aligned}$$

#### 4.4 ESTRATEGIA DE MUESTREO

Si se realiza un análisis de cada uno de los  $h$  incrementos de muestra (véase el ejemplo anterior, Sec. 4.3) entonces los límites de confianza de la media están dados por la ecuación (2.9):

$$\mu = \bar{x} \pm t(s/\sqrt{h})$$

donde  $\bar{x}$  es la media de las mediciones y  $s^2$  la varianza de ellas;  $s^2$  es una estimación de la varianza total,  $\sigma^2$ , que es la suma de las varianzas muestrales y de mediciones, es decir,  $\sigma_0^2 + \sigma_1^2$  (véase la Sec. 1.6 y 2.7), y  $\sigma^2/h$  (que es estimada por  $s^2/h$ ) es la varianza de la media,  $\bar{x}$ . Si el valor de cada incremento de muestra es la media de  $n$  mediciones repetidas, entonces la varianza de la media se convierte en  $(\sigma_0^2/n + \sigma_1^2)/h = \sigma_0^2/nh + \sigma_1^2/h$ . Es obvio que para alcanzar la precisión máxima se requiere que la varianza de la media sea tan pequeña como fuera

posible. El término debido a la varianza de la medición puede ser reducido ya sea utilizando un método de análisis más preciso o incrementando  $n$ , el número de mediciones repetidas; pero no hay que esforzarse por hacer que la varianza de la medición sea menor que un décimo de la varianza muestral, ya que cualquier reducción posterior no mejorará en absoluto la varianza total (que es la suma de las dos). Es preferible, en cambio, tomar un número mayor de incrementos de muestra, ya que el intervalo de confianza disminuye cuando aumenta  $h$ . Si se ha utilizado una muestra preliminar para estimar  $s$ , entonces el tamaño de la muestra requerido para lograr un tamaño dado del intervalo de confianza se puede calcular de manera aproximada (véase el Cap. 2, Ej. 4).

Una estrategia de muestreo con materiales a granel es tomar  $h$  incrementos de muestra y mezclarlos antes de hacer  $n$  mediciones repetidas. La varianza de la media de éstas mediciones es:  $\sigma_0^2/n + \sigma_1^2/h$ . Esta varianza total debería compararse con la resultante de analizar  $n$  veces cada incremento de muestra y promediar las medias de los incrementos, siendo entonces la varianza  $\sigma_0^2/nh + \sigma_1^2/h$  (véase lo anterior). Obviamente la última varianza es la más pequeña, dando por resultado una media con más precisión, aunque se requieren más mediciones ( $nh$  frente a  $n$ ). El conocimiento de los valores  $\sigma_0^2$  y  $\sigma_1^2$  de experiencias anteriores, y los costos de muestreo y análisis, pueden utilizarse para calcular el costo de las estrategias de muestreo relativas. Por lo general se usará el esquema más económico que proporcione el grado de precisión requerido.

En el caso de los materiales a granel la varianza muestral depende del tamaño del incremento de muestra relativo a la escala de no homogeneidades, y disminuye con el aumento del tamaño del incremento de muestra. En algunas experiencias puede ser necesario seleccionar un límite superior de la varianza muestral, de manera que puedan detectarse cambios en la media. Pueden tomarse mediciones preliminares para decidir el tamaño del incremento de muestra mínimo para proporcionar un nivel aceptable de varianza muestral.

#### 4.5 ENSAYOS DE COLABORACIÓN: INTRODUCCIÓN

Los ensayos de colaboración pretenden dar respuesta a la siguiente pregunta: "¿Si la misma muestra (o una serie de partes alicuotas idénticas de una única muestra) se analiza por el mismo método en diferentes laboratorios, los resultados obtenidos serán los mismos dentro de los límites de error aleatorio? Muy a menudo se supone que, con el uso de instrumentos idénticos en diferentes laboratorios, se obtendrán resultados muy similares cuando se aplica un método particular a muestras idénticas por científicos expertos. Por desgracia, la amplia experiencia práctica, ha mostrado en repetidas ocasiones que esto no se cumple y que incluso pueden obtenerse resultados muy divergentes aún cuando se piense que se han tomado todas las precauciones experimentales. Por ejemplo, en un estudio del nivel de ácidos grasos polinsaturados en una muestra de aceite de palma, realizado en 16 laboratorios diferentes, éste varió entre 5.5% y 15.0%! La gran variación fue muy notoria por el hecho de que no fue un análisis de trazas y de que se utilizó un método para determinar ácidos grasos relativamente sencillo por espectrometría de absorción, después de saponificar y digerir con una enzima la muestra original. En otra situación, al determinar el porcentaje de aluminio en una muestra de piedra caliza, en diez laboratorios, se obtuvieron valores en el intervalo del 1.11% al 1.9%.

Estos resultados divergentes deben considerarse con extrema cautela; implican que, por ejemplo, una muestra de un alimento que al parecer supera una prueba de seguridad en un laboratorio, es probable que no la supere si se examina en otro. El suero sanguíneo de un individuo puede parecer completamente normal en un laboratorio y, sin embargo, indicar un estado patológico en otro. Tales posibilidades son claramente inaceptables: los métodos analíticos que se utilizan en laboratorios diferentes tienen que elegirse con cuidado a fin de minimizar las variaciones entre laboratorios y los consiguientes problemas de interpretación de sus resultados. Generalmente sólo se adoptarán los métodos que proporcionen pequeñas variaciones entre los laboratorios para su uso en análisis estándar y confiables. Se ha prestado mucha atención al diseño de ensayos de colaboración, de manera que, con el mínimo esfuerzo, produzcan la máxima cantidad de información sobre las fuentes y la magnitud de los errores entre los laboratorios.

#### 4.6 REPRESENTACIONES GRÁFICAS DE DOS MUESTRAS

La alarmante dispersión de resultados obtenidos en diferentes laboratorios durante ensayos de colaboración se podría explicar, en principio, por los *errores aleatorios* de las mediciones, o por la presencia de diferentes *errores sistemáticos* en diferentes laboratorios. En la práctica, la evidencia es abrumadora en cuanto a que las variaciones son causadas por errores sistemáticos. Este resultado se demuestra mucho mejor usando un diagrama para dos muestras (o  $X - Y$ ), ideado por W. J. Youden. El principio en que se basa es que a cada laboratorio que colabora en el ensayo se le debe enviar *dos muestras similares* ( $X$  y  $Y$ ), y se le solicita que realice una determinación de cada una. Los resultados se grafican como se muestra en la figura 4.1, donde cada punto es un par de resultados de un único

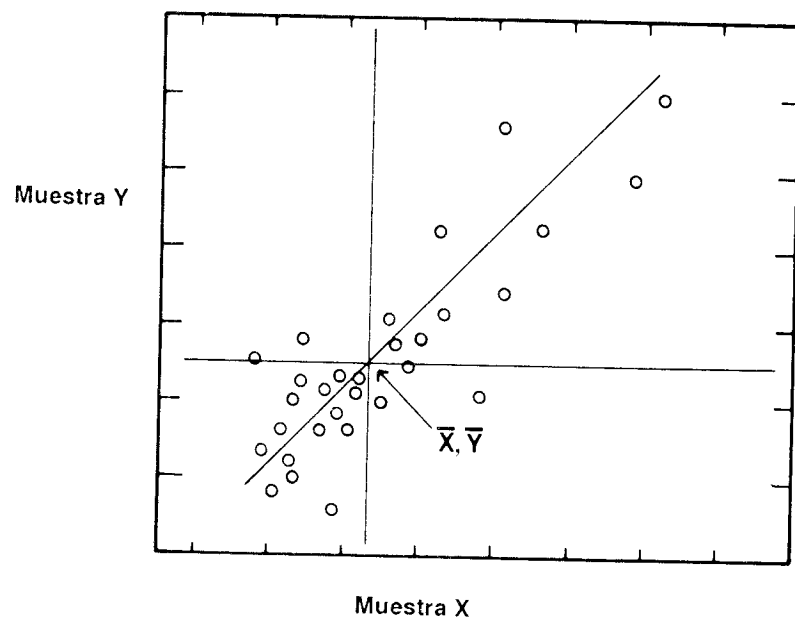


Figura 4.1 Una representación gráfica de dos muestras de Youden.

laboratorio. También se determinan los valores medios de  $X$  y  $Y$  y se dibujan líneas horizontales y verticales a través del punto  $(\bar{X}, \bar{Y})$ , con lo cual se divide el diagrama en cuatro cuadrantes.

Si los resultados de los diferentes laboratorios varían por completo debido a errores aleatorios, las determinaciones de  $X$  y  $Y$  en cada caso pueden dar resultados que son a la vez demasiado altos, o demasiado bajos,  $X$  alto y  $Y$  bajo, o  $X$  bajo y  $Y$  alto. Estos cuatro resultados serían igualmente probables y darían aproximadamente el mismo número de puntos en cada uno de los cuatro cuadrantes del diagrama. Pero si los errores sistemáticos son la causa predominante de la variación, podremos esperar que un laboratorio que obtiene un valor alto de  $X$  tiende también a obtener un valor alto de  $Y$ . Esto conduciría a un predominio de puntos en los cuadrantes superior derecho e inferior izquierdo del diagrama, y éste es de hecho el resultado obtenido en prácticamente todos los ensayos de colaboración. Es obvio que, ante la hipotética ausencia de errores aleatorios, todos los puntos deberían caer sobre la diagonal de 45 del diagrama. En la práctica, ya que en alguna medida los errores aleatorios están siempre presentes, los puntos caen dentro de una elipse la cual, tiene esta diagonal como su eje mayor. La longitud de la perpendicular desde un punto individual a la diagonal da una medida del error aleatorio, y la perpendicular interseca a la diagonal en un punto a una distancia de  $(\bar{X}, \bar{Y})$  que está relacionado con el error sistemático de ese laboratorio. Es evidente que esta aproximación de dos muestras a los ensayos de colaboración puede proporcionar mucha información útil. Youden ha señalado otras ventajas de esta aproximación, las cuales se analizarán en la sección 4.8.

#### 4.7 LA PREPARACIÓN DE UN ENSAYO DE COLABORACIÓN

El propósito de un ensayo de colaboración es claro: evaluar la variación en los resultados de un método analítico concreto (a menudo uno nuevo o modificado) cuando se efectúa en una serie de laboratorios. Los resultados se podrían utilizar para decidir si el método mismo es apto para uso general, y quizás también para identificar aquellos laboratorios a los que se les podría confiar un trabajo analítico importante, por ejemplo, con repercusiones para la salud pública o de otro tipo de normatividad. De esta manera, al parecer es fácil establecer este ensayo. Un único laboratorio "de arbitraje" (a menudo, en la práctica, un laboratorio del gobierno) debería enviar una o más muestras a otros laboratorios, junto con instrucciones detalladas sobre los métodos analíticos que se deban utilizar. Los laboratorios de colaboración analizarán las muestras de acuerdo con estas instrucciones, y devolverán sus resultados al centro de arbitraje, en donde se evaluarán utilizando métodos estadísticos. En la práctica, se emplea una buena cantidad de trabajo para establecer y llevar a cabo dicho ensayo, y se deben tomar decisiones muy importantes antes de enviar las muestras.

Un experimento preliminar decisivo es la prueba de "entereza". Como observamos en el capítulo 1, incluso los experimentos analíticos muy sencillos suponen varios pasos individuales y, quizás, el uso de un cierto número de reactivos. Así, muchos factores experimentales (por ejemplo, temperatura, composición del disolvente, pH, humedad, pureza de reactivos y concentración, etc.) afectarán los resultados, y es esencial que dichos factores se identifiquen y estudien antes de emprender cualquier ensayo de colaboración. En algunos casos, un método resulta ser tan sensible a pequeños cambios en un factor, que en la práctica es muy difícil de controlar (por ejemplo, reactivos de alta pureza), de manera que se rechaza por ser impracticable antes de que se efectúe el ensayo. En otros casos el

ensayo continuará, pero se prevendrá a los colaboradores de aquellos factores que deben ser controlados con cuidado. Aunque en el capítulo 7 se expone un análisis más completo del diseño experimental, es importante indicar aquí que puede obtenerse mucha información de un número relativamente pequeños de experimentos. Suponga que se cree que siete factores experimentales (A a G) podrían afectar los resultados de un análisis. Estos factores tienen que ser probados con (al menos) dos valores, denominados **niveles**, para ver si son realmente significativos. Así, si se pensó que la temperatura afectaba el resultado, debemos realizar experimentos preliminares a dos temperaturas (niveles) y comparar los resultados. De manera similar, si la pureza de los reactivos puede ser importante, se deben realizar experimentos con lotes de reactivos de alta y baja pureza. Podría pensarse que serán necesarios  $2^7$  experimentos preliminares para cubrir todas las combinaciones posibles de siete factores a dos niveles. Sin embargo, en la práctica sólo ocho experimentos proporcionarán información importante. Los dos niveles de los factores se denominan + y -, y la tabla 4.2 muestra cómo están situados en los ocho experimentos, cuyos resultados se denominan  $y_1, y_2, \dots, y_8$ . El efecto de alterar cada uno de los factores de su nivel alto al bajo se calcula fácilmente. Así el efecto de cambiar B de + a - está dado por:  $(y_1 + y_2 + y_5 + y_6)/4 - (y_3 + y_4 + y_7 + y_8)/4$ . Cuando las siete diferencias de los factores A a G hayan sido calculados en su totalidad de este modo, es fácil identificar cualquier factor que tenga un efecto preocupante en los resultados. Se puede demostrar que cualquier diferencia que sea más de dos veces la desviación estándar de mediciones repetidas es significativa y debería estudiarse posteriormente. Este simple conjunto de experimentos, conocido técnicamente como **diseño factorial incompleto**, tiene la desventaja de que no se pueden detectar las interacciones entre los factores. Este tema se analizará en el capítulo 7.

Tabla 4.2 Prueba de entereza para siete factores

Experimento	Factores							Resultado
	A	B	C	D	E	F	G	
1	+	+	+	+	+	+	+	$y_1$
2	+	+	-	+	-	-	-	$y_2$
3	+	-	+	-	+	-	-	$y_3$
4	+	-	-	-	-	+	+	$y_4$
5	-	+	+	-	-	+	-	$y_5$
	-	+	-	-	+	-	+	$y_6$
	-	-	+	+	-	-	+	$y_7$
8	-	-	-	+	+	+	-	$y_8$

Otro aspecto que se debe establecer antes de comenzar el ensayo es el número de laboratorios participantes. Por supuesto éste no debería ser demasiado pequeño, ya que los resultados tendrían escasa aplicación general. Puesto que las diferencias sistemáticas entre laboratorios causarían muchas de las variaciones en los resultados, es más adecuado enviar unas pocas muestras a numerosos laboratorios que enviar muchas muestras a unos pocos, aunque utilizar un gran número de laboratorios será costoso y empleará tiempo. En la práctica, muchos ensayos de colaboración se realizan en forma satisfactoria con 10 a 20 laboratorios. Si el centro de arbitraje conoce a los participantes, puede ser correcto seleccionar deliberadamente un intervalo de laboratorios variando en calidad de excelente

a insatisfactorio. Otra alternativa es numerar todos los laboratorios disponibles y seleccionar a los participantes, mediante una tabla de números aleatorios (véase la Sec. 4.2). Por supuesto, no se permite seleccionar sólo laboratorios de clase alta.

La preparación, el empaquetado y el envío de las muestras es otro elemento importante del ensayo. Es normal tener una reserva de muestra y dividirla entre los laboratorios participantes (el arbitro retiene algunas muestras para permitir cualquier experimento adicional necesario). Si la muestra no es homogénea por naturaleza, pueden surgir los problemas de muestreo analizados en las secciones precedentes.

4.8 CÁLCULOS EN ENSAYOS DE COLABORACIÓN

Cuando se concluye el trabajo experimental de un ensayo de colaboración, y se envían los datos al organismo de arbitraje, debe realizarse la evaluación estadística de los resultados. Estos cálculos deberían revelar hasta qué punto la variación entre los laboratorios participantes se debe a errores sistemáticos. En esta sección se consideran los cálculos inherentes a dos casos muy simples de ensayos de colaboración, y también algunos de los problemas que pueden surgir en la práctica.

Quizás el caso más sencillo de ensayo sea aquel en el que a cada laboratorio participante se le envía una única muestra y sobre ella se realizan mediciones repetidas. Deben analizarse los resultados enviados al organismo de arbitraje para separar la variación entre las repeticiones de la variación entre laboratorios. El siguiente ejemplo muestra que esto se puede hacer mediante una aplicación simple del ANOVA de un factor: en este caso, tenemos un factor de efecto aleatorio, como se describió en la sección 4.3.

Ejemplo: se envió una muestra de pienso medicinal que contiene el factor de crecimiento olaquinox a cinco laboratorios, cada uno de los cuales realizó análisis por duplicado utilizando cromatografía líquida de alta resolución con un detector ultravioleta. Se obtuvieron los siguientes resultados:

Laboratorio	Olaquinox encontrado (mg/kg)	
1	23.0	22.2
2	23.8	22.6
3	21.0	23.8
4	26.5	27.1
5	21.4	21.4

(Datos tomados de Analytical Methods Committee, *Analyst*, 1985, 110, 75).

Estime las variaciones aleatorias y entre laboratorios en este ensayo.

Los cálculos ANOVA comienzan restando 20 a cada resultado, lo cual da:

Laboratorio	$T_i$		$T_i^2$
1	3.0	2.2	5.2
2	3.8	3.6	7.4
3	1.0	3.8	4.8
4	6.5	7.1	13.6
5	1.4	1.4	2.8
$T = 33.8$			$\sum T_i^2 = 297.64$

$n = 2 \quad h = 5 \quad N = 10 \quad \sum_i \sum_j x_{ij}^2 = 153.26$

La tabla ANOVA es, entonces:

Fuente de variación	Suma de cuadrados		Grados de libertad	Cuadrado medio
Entre laboratorios	$(297.64/2) - (33.8)^2/10$	$= 34.576$	4	8.644
Dentro del laboratorio	por diferencias	$= 4.440$	5	0.888
Total	$153.26 - (33.8)^2/10$	$= 39.016$	9	

Ya que el cuadrado medio entre laboratorios es mayor que el de dentro de los laboratorios, se utiliza la prueba  $F$  para ver si esta diferencia es significativa. El valor de  $F$  en este caso está dado por:

$$F_{4,5} = 8.644/0.888 = 9.734$$

El valor crítico de  $F_{4,5}$  ( $P = 0.05$ , tabla A.2) es 5.192, por lo que se concluye que la diferencia entre los dos cuadros medios es significativa. Esto indica que la variación sistemática entre laboratorios  $\sigma_1^2$  se encuentra muy por encima de cero, y puede estimarse como [(cuadrado medio entre muestras) - (cuadrado medio dentro de las muestras)]/ $n$  (véase la Sec. 4.3). El valor resultante para  $\sigma_1^2$  es 3.878 mostrando claramente que la gran parte de la variación de los resultados se debe por supuesto a diferencias sistemáticas entre laboratorios.

En la sección 4.6 se mostró que un ensayo en el que *cada laboratorio realiza una sola medición sobre cada una de dos muestras similares*, proporciona información valiosa acerca de los errores aleatorios y los sistemáticos. Esta aproximación tiene la ventaja adicional de que los laboratorios participantes no tienden a suprimir una o más determinaciones repetidas, y que se pueden estudiar muchos materiales sin considerar un gran número de experimentos.

*Ejemplo.* Se determinaron en nueve laboratorios (1-9) los niveles de plomo (en ng/g) en dos muestras similares ( $A$  y  $B$ ) de fórmulas de leche en polvo para niños, utilizando espectrometría de absorción atómica con horno de grafito. Los resultados fueron:

Muestra	Laboratorios								
	1	2	3	4	5	6	7	8	9
$A$	35.1	23.0	23.8	25.6	23.7	21.0	23.0	26.5	21.4
$B$	33.0	23.2	22.3	24.1	23.6	23.1	21.0	25.6	25.0

Estime la variación global entre laboratorios y sus componentes aleatorios y sistemáticos.

Este ejemplo difiere del anterior en que existen diferencias entre las muestras así como entre los laboratorios. Por el procedimiento normal, tal situación debería ser tratada por un ANOVA de dos factores (véase la Sec. 7.7). Sin embargo, en este ejemplo hay solamente *dos* muestras, elegidas deliberadamente con un contenido analítico similar, de manera que no existe interés en evaluar la diferencia entre ellas. Además, se puede hacer los cálculos por un camino que es numérica y conceptualmente más simple que un ANOVA de dos factores. Al realizar el cálculo observamos que el

resultado obtenido por cada laboratorio para la muestra  $A$  puede incluir un error sistemático. El mismo error se incluirá presumiblemente en aquel resultado de laboratorio para la muestra similar  $B$ . Por lo tanto, la diferencia  $D$  ( $= A - B$ ) tendrá eliminado este error, de manera que la diversidad de los valores de  $D$  proporcionará una estimación de los errores aleatorios o de la medición. En forma similar  $A$  y  $B$  pueden sumarse para dar  $T$ , cuya diversidad proporciona una estimación de la variación global en los resultados. La varianza de medición se estima entonces por:

$$\sigma_0^2 = \frac{\sum(D_i - D)^2}{2(n-1)} \quad (4.1)$$

y la varianza global,  $\sigma^2$ , debida a todas las fuentes de error, se estima por:

$$\sigma^2 = \frac{\sum(T_i - T)^2}{2(n-1)} \quad (4.2)$$

Advierta que cada una de estas ecuaciones incluye un 2 en el denominador. Esto se justifica porque tanto  $D$  como  $T$  dan estimaciones de los errores en los *dos* conjuntos de resultados, los cuales se restan y suman a  $D$  y  $T$  respectivamente. Los resultados de este ensayo se pueden expresar en una tabla como la siguiente:

	1	2	3	4	5	6	7	8	9
$A$	35.1	23.0	23.8	25.6	23.7	21.0	23.0	26.5	21.4
$B$	33.0	23.2	22.3	24.1	23.6	23.1	21.0	25.6	25.0
$D$	2.1	-0.2	1.5	1.5	0.1	-2.1	2.0	0.9	-3.6
$T$	68.1	46.2	46.1	49.7	47.3	44.1	44.0	52.1	46.4

Otros cálculos obtenidos a partir de la tercera y cuarta filas de datos muestran que  $D = 0.244$ , y  $T = 49.33$ , y de aquí que la estimación de  $\sigma_0^2$  sea  $(1.383)^2$ , y la estimación de  $\sigma^2$  sea  $(5.296)^2$ . Por tanto,  $F_{8,8} = (5.296/1.383)^2 = 14.67$ . El valor crítico ( $P = 0.05$ ) es 3.44 (tabla A.2). El valor experimental lo supera ampliamente, de manera que se concluye que  $\sigma^2$  es de hecho mucho más grande que  $\sigma_0^2$ . Esto implica, como hemos visto, que la variación entre laboratorios no puede explicarse sólo por errores de medición aleatorios, y que se han producido errores sistemáticos entre laboratorios. La componente de  $\sigma^2$  debida a estos errores sistemáticos,  $\sigma_1^2$ , se calcula a partir de:

$$\sigma^2 = 2\sigma_1^2 + \sigma_0^2 \quad (4.3)$$

Note de nuevo la aparición del 2 en la ecuación porque se estudian dos muestras. Aquí es fácil calcular que la estimación de  $\sigma_1^2$  es  $(3.615)^2$ .

Aunque los ensayos de colaboración de este tipo parecen muy simples tienen también sus problemas. A veces ocurre que un laboratorio no puede tomar las medidas sobre las dos o sobre todas las muestras que le mandan, quizás porque se haya perdido o dañado una muestra durante el envío, o a causa de errores accidentales en el propio laboratorio. La desaparición de datos presenta problemas de interpretación que se tratan con profundidad en textos más avanzados.

Otro problema avanzado en ensayos de colaboración es el de los valores anómalos; es decir, uno o más laboratorios pueden generar resultados que al parecer se encuentran fuera de lugar con respecto al resto. En el segundo ejemplo anterior, resulta obvio que los resultados obtenidos por el laboratorio 1 son considerablemente más altos que los otros datos, y la aplicación de la prueba  $Q$  (Sec. 3.7) muestra que estos resultados podrían ser rechazados al nivel de probabilidad  $P = 0.05$ . Si se vuelven a calcular los valores de  $\sigma_0^2$  y  $\sigma^2$  para los laboratorios restantes se demostraría que, aunque el último valor sigue siendo mayor, la prueba  $F$  indica que la diferencia ya no es significativa ( $P = 0.05$ ). Esto es, si se eliminase el laboratorio 1 del ensayo, toda la variación restante podría deberse sólo a errores de medición aleatorios. En la práctica, el organismo de arbitraje intentará establecer si existe alguna razón concreta en cuanto a los resultados de este laboratorio. Si se dispusiera de tal explicación, podría plantearse la posibilidad de rechazar los datos.

Muchos ensayos de colaboración son bastante más complejos que los vistos hasta ahora. Por lo regular, se suelen enviar varias muestras a cada uno de los laboratorios participantes (entre 10 y 20), cada uno de los cuales tomaría dos o más mediciones repetidas sobre cada muestra. Al efectuar la evaluación estadística de los resultados de tales ensayos se debe tener en cuenta tanto la variación entre las muestras como los errores de medida aleatorios y las variaciones entre laboratorios. Dichos cálculos se examinan posteriormente en el capítulo 7.

#### 4.9 DIAGRAMAS DE CONTROL

Con mucha frecuencia surge la necesidad de controlar periódicamente la fabricación de un producto, quizás para observar si los artículos individuales contienen, en promedio, la cantidad correcta de una sustancia dada, y para asegurar que no existe demasiada variación. Una forma de hacerlo es tomando pequeñas muestras a intervalos regulares. Considere la situación concreta en que se controla el peso de las pastillas que salen de una línea de producción. Idealmente la medida del peso de cada pastilla debería ajustarse a un valor exacto,  $\mu_0$ , aunque en la práctica habrá cierta variación aleatoria de una pastilla a otra. Esta variación se debe en parte al error aleatorio al realizar la medición y en parte a las diferencias reales en peso entre las pastillas. Suponga que conocemos el tamaño de la variación aleatoria global, medida por la desviación estándar de la población,  $\sigma$ , de experiencias anteriores; si el proceso está bajo control, es decir que, los pesos de las pastillas producidas *si* tienen el peso medio  $\mu_0$ , y desviación estándar  $\sigma$ , entonces para muestras de  $n$  pastillas, cerca del 95% de las medias muestrales se encontrarán incluidas en el intervalo  $\mu_0 \pm 2\sigma/\sqrt{n}$  y aproximadamente el 99.7% dentro del intervalo  $\mu_0 \pm 3\sigma/\sqrt{n}$  [véase Cap. 2, Ecs. (2.5) y (2.7)]. Un **diagrama de control** es una gráfica en la que se presentan consecutivamente los valores de las medias muestrales de manera que se pueda tomar cualquier acción correctora lo más rápido posible. En la figura 4.2 se muestra un tipo de diagrama de control, conocido como **diagrama de Shewhart**, para el proceso descrito. Se representa el valor de la media muestral,  $\bar{x}$ , en función tiempo, y cuando el proceso está bajo control los valores de  $\bar{x}$  se distribuyen normalmente en torno a  $\mu_0$ . Hay también dos pares de líneas horizontales en el diagrama: las **líneas de alerta** en  $\mu_0 \pm 2\sigma/\sqrt{n}$  y las **líneas de acción** en  $\mu_0 \pm 3\sigma/\sqrt{n}$ . El objeto de estas líneas está indicado por sus propios nombres. Puesto que la probabilidad de que una media caiga fuera de las líneas de acción, cuando el proceso está bajo control, es solamente 0.003 (es decir, 3 de

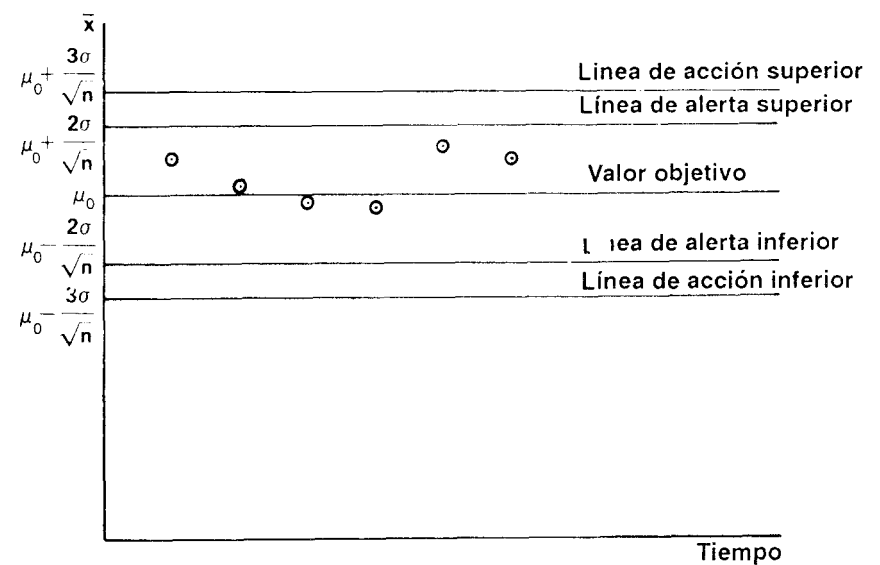


Figura 4.2 Un diagrama de Shewhart típico.

cada 1000), en el caso de que esto ocurra lo normal será detener el proceso e investigar lo sucedido. La probabilidad de que  $\bar{x}$  caiga entre cada par de líneas de acción y de alerta está próxima a 0.025 (es decir, 1 en 40): un punto que cayera aquí no sería causa de preocupación, pero si dos puntos consecutivos caen en esta zona habría que detener el proceso. Se puede observar que un diagrama de control muestra una serie de pruebas de significación, correspondiendo las líneas de aviso y de acción a los valores críticos para  $P = 0.05$  y  $P = 0.003$  respectivamente. En ocasiones el proceso se podría detener cuando está todavía bajo control (un error de Tipo 1, véase la Sec. 3.13), aunque, el riesgo de que esto suceda es bajo.

Existen dos razones por las que los puntos podrían empezar a caer fuera de las líneas de acción. La primera, la media de los artículos producidos conocida como media del proceso, puede haber derivado hacia fuera del valor objetivo; en segundo lugar puede existir un incremento en la variación, de manera que las líneas de aviso y de acción no indiquen las probabilidades correctas, sino que estén demasiado próximas al valor objetivo. Recíprocamente, un *descenso* en la variación significa que puede no detectarse un cambio en la media del proceso, debido a que las líneas de aviso y de acción no indican las probabilidades correctas y están demasiado lejos del valor objetivo. Por estas razones es importante estudiar la variación de un proceso así como su valor medio. Esto se hace con mayor facilidad utilizando el **intervalo**,  $w$  (es decir, la diferencia entre el valor más grande y el más pequeño) de cada muestra tomada. El diagrama de Shewhart adecuado se puede construir con la ayuda de tablas que proporcionan las líneas de aviso y de acción, y el valor objetivo de  $w$  para valores diferentes de  $n$  y  $\sigma$  (véase la bibliografía de este capítulo). Los diagramas de medias e intervalos se utilizan simultáneamente, y se toman acciones apropiadas si los puntos caen fuera de las líneas sobre cualquier diagrama.

Todavía no hemos considerado con detalle cómo establecer un valor para  $\sigma$ , esto es decisivo al construir los diagramas de control para  $\bar{x}$  y  $w$ ; a veces  $\sigma$  se denomina **capacidad**



del proceso, debido a que mide la variación inherente. Ya que es una desviación estándar de una población debe estimarse a partir de una muestra muy grande, por ejemplo, de pastillas. Esto podría causar dificultades: necesitamos estar seguros de que cuando se tomen estas medidas no habrá derivas en la media del proceso que proporcionarían un valor erróneamente alto de  $\sigma$ , ¡pero no hemos construido todavía el diagrama para seguir la media! Este problema circular se supera tomando un gran número de pequeñas muestras durante un periodo. A partir de cada muestra se puede calcular una estimación de la varianza dentro de muestras que no se ve afectada por las posibles variaciones entre las muestras. Estas estimaciones se pueden promediar para dar un valor de la capacidad del proceso [véase la Sec. 3.9 (i)]. En la práctica, por lo general se utilizan los intervalos de las muestras, ya que esto conduce a cálculos más simples, aunque implica el mismo principio.

Una consideración importante al utilizar un diagrama de Shewhart es la rapidez con que se detecta un cambio brusco en la media del proceso. Cuando ésta cambia, el siguiente punto no caerá necesariamente fuera de las líneas de aviso o de acción; por ejemplo, si cambia en  $3\sigma/\sqrt{n}$ , entonces existe una probabilidad de 1/2 de que el siguiente punto caiga fuera de las líneas de acción, pero si la media del proceso cambia en  $1\sigma/\sqrt{n}$ , esta probabilidad se reduce a 1/40. El número promedio de puntos que se deben representar antes de que sea detectado un cambio en la media del proceso se denomina **longitud media de racha (LMR)**. Por supuesto, ello depende del tamaño del cambio en la media del proceso en comparación con  $\sigma/\sqrt{n}$ : cuanto mayor sea el cambio, se detectará con más rapidez. Por ejemplo, si la media del proceso cambia en  $1\sigma/\sqrt{n}$ , entonces la LMR antes de que una media muestral caiga fuera de las líneas de acción es aproximadamente de 50. Si el proceso se detiene cuando dos puntos consecutivos caen fuera de la línea de alerta, la LMR queda aproximadamente dividida por dos. Puesto que la LMR depende del cambio en la media del proceso comparada con  $\sigma/\sqrt{n}$ , se puede reducir aumentando el tamaño de la muestra: el mayor tamaño que puede tomarse en la práctica suele estar dado por consideraciones de costos y tiempos: por lo regular se toman tamaños de muestra entre 4 y 6.

Un tipo diferente de diagrama de control, conocido como **diagrama de sumas acumuladas (DSA)**, utiliza todas las medias muestrales previas, en lugar de la última o de las dos últimas como hace el diagrama de Shewhart. SA son las siglas del término “suma acumulada”, es decir, la suma de las desviaciones de las medias muestrales del valor objetivo, acumuladas de atrás hacia adelante. Un ejemplo aclarará el cálculo:

*Ejemplo.* En la tabla 4.3 se proporciona el valor de la media para una serie de muestras. Se sabe que el valor de  $\sigma/\sqrt{n}$  es 2.5 y la figura 4.3 muestra el diagrama de Shewhart para las medias naturales. Se ve que, aunque no quedan puntos fuera de las líneas de aviso, una serie de valores cae a un lado del valor objetivo.

En la tabla 4.3 también se muestra el cálculo de las sumas acumuladas. Es obvio que si el proceso está bajo control, las desviaciones positivas y negativas son igualmente probables, por lo que la suma acumulada debería oscilar alrededor de cero; estos valores se encuentran representados en la figura 4.4. Para ofrecer una buena impresión visual, el DSA aparece dibujado de tal forma que la distancia correspondiente a  $2\sigma/\sqrt{n}$  sobre el eje vertical es igual a la distancia entre observaciones sucesivas sobre el eje horizontal.

Del DSA se desprende que la media del proceso cambió después de tomar la octava muestra. Una ventaja del DSA es que indica en qué punto el proceso se salió de control.

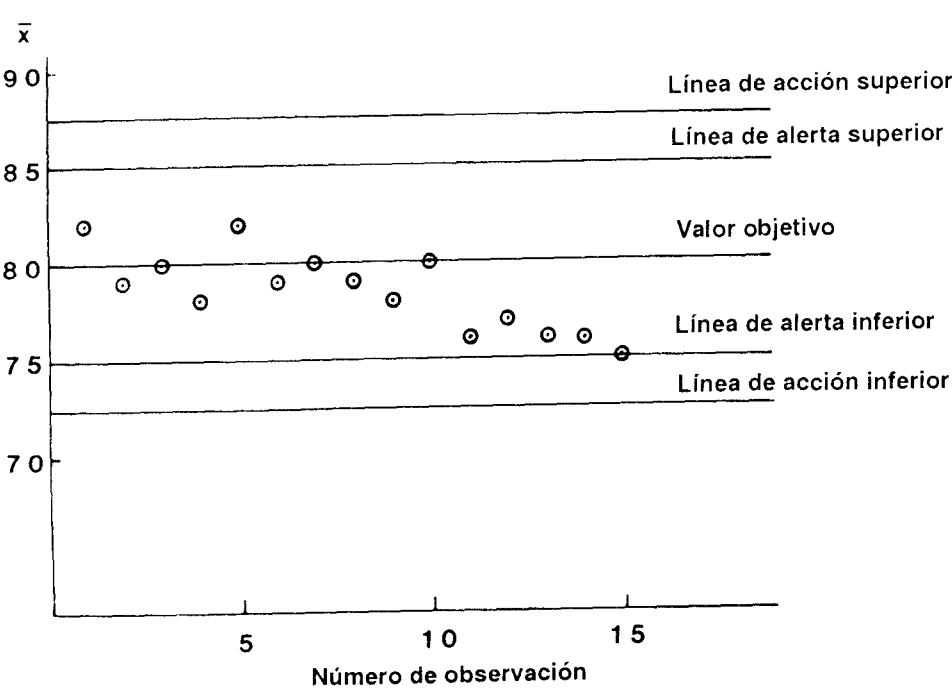


Figura 4.3 Diagrama de Shewhart para los datos de la tabla 4.3.

Tabla 4.3 Cálculo de la suma acumulada: valor objetivo = 80;  $\sigma/\sqrt{n}$  (conocido) = 2.5

Número de observación	Media muestral	(Media muestral) - (valor objetivo)	SA
1	82	+2	+2
2	79	-1	+1
3	80	0	+1
4	78	-2	-1
5	82	+2	+1
6	79	-1	0
7	80	0	0
8	79	-1	-1
9	78	-2	-3
10	80	0	-3
11	76	-4	-7
12	77	-3	-10
13	76	-4	-14
14	76	-4	-18
15	75	-5	-23

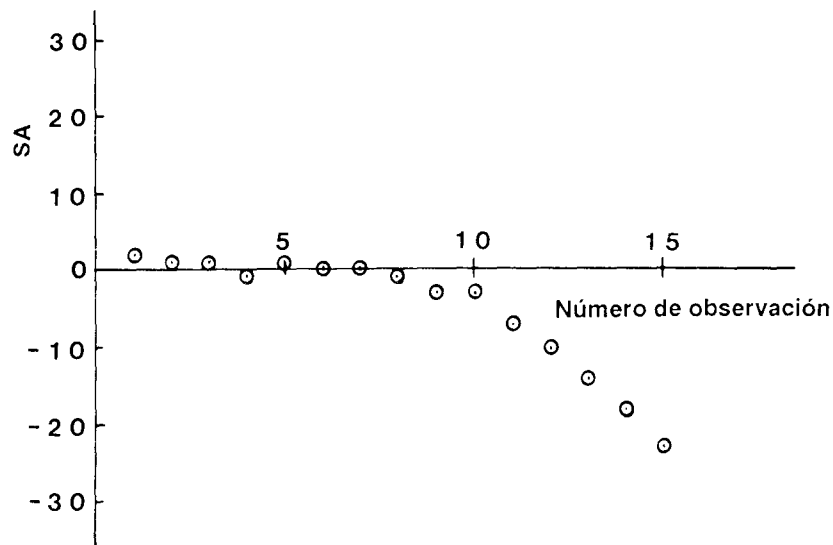


Figura 4.4 Diagrama de sumas acumuladas (SA) para los datos de la tabla 4.3.

Para probar si una tendencia en un DSA *indica* que la media del proceso ha cambiado y que esto no se puede atribuir a simples variaciones aleatorias, puede utilizarse un **delimitador V**. Como se muestra en la figura 4.5 un delimitador en forma de V, de preferencia grabado en plástico claro para poder superponerse (transparencias convencionales), se sitúa en el DSA con su eje horizontal y su vértice a una distancia  $d$  a la derecha de la última observación. El semiángulo entre los brazos de la V se denomina  $\theta$ . Se dice que el proceso está bajo control si todos los valores de la SA caen entre los brazos de la V, como se indica en la figura 4.5. La figura 4.6 ilustra una situación en la que el proceso está fuera de control: dos de los valores de SA se encuentran fuera del brazo superior de la V, lo que indica que la media del proceso ha caído por debajo del valor objetivo (véanse las Figs. 4.3 y 4.4). Por supuesto, el rendimiento del delimitador dependerá de los valores elegidos para  $\theta$  y  $d$ . Cuanto mayor sea el valor de  $d$  (y menor el valor de  $\theta$ ), más probable será que un valor dado caiga fuera de los brazos de la V. Estos valores deben elegirse de manera que aparezcan pocas falsas alarmas cuando el proceso está bajo control pero que se detecta rápidamente un cambio importante en la media del proceso. Otros detalles acerca de la dependencia de la LMR de  $\theta$  y  $d$  para diferentes cambios en la media del proceso se encuentran en una monografía I.C.I. citada en la bibliografía de este capítulo. En general, la LMR para un DSA es inferior que para un diagrama de Shewhart con el mismo tamaño de muestra.

Un DSA también puede utilizarse para estimar el tamaño del cambio ocurrido en la media del proceso cuando éste se encuentra fuera de control. Si, por ejemplo, la media del proceso disminuye en  $\Delta$ , entonces cada media muestral, en promedio, será el valor objetivo menos  $\Delta$ . Como resultado, la SA disminuirá en promedio en  $\Delta$  para cada punto representado. Así pues, la pendiente promedio de la línea que une los puntos de la SA proporciona una medida del cambio en la media del proceso y, por consiguiente, la corrección necesaria.

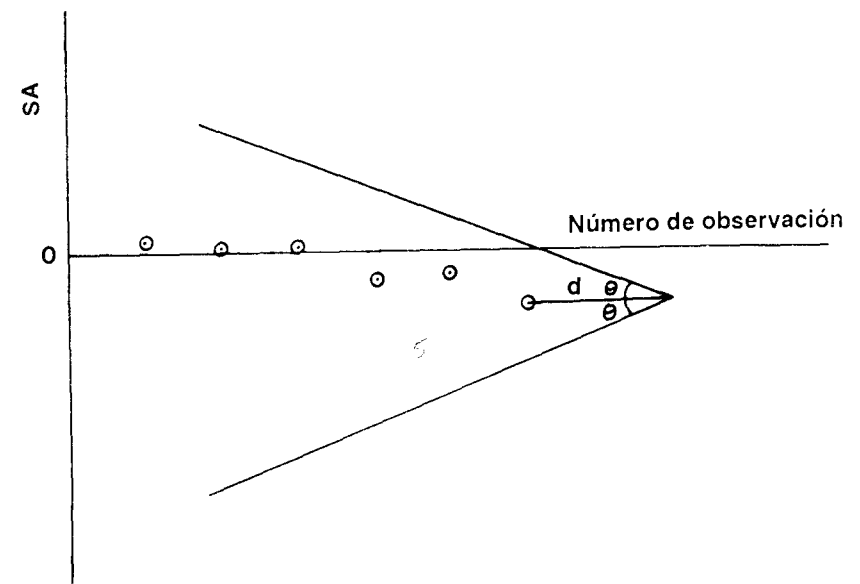


Figura 4.5 Uso de un delimitador V con el proceso bajo control.

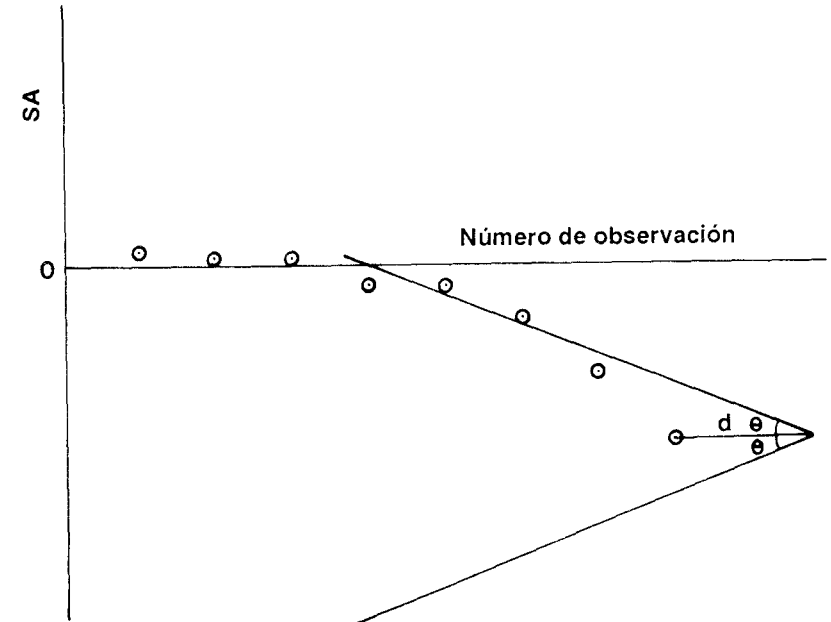


Figura 4.6 Uso de un delimitador V con el proceso fuera de control.

Los métodos descritos en esta sección se pueden utilizar también para controlar la exactitud y precisión de los análisis de rutina en el laboratorio. A intervalos de tiempo fijos sobre una muestra estándar, se toma un número pequeño de muestras repetidas, y la media y el intervalo de las mediciones repetidas se representan sobre los diagramas de control. El diagrama de control para la media controla la exactitud, y la concentración conocida de analito en el estándar es el valor objetivo. Para el intervalo, esta consideración controla la precisión, y el valor objetivo es la capacidad del proceso. De nuevo es necesario establecer un valor para la capacidad del proceso con el fin de construir los diagramas de control. En este caso, la capacidad del proceso estará limitada por el error aleatorio que se produce al tomar las medidas, en lugar del error que se comete al preparar un estándar. Esto se puede verificar al preparar una serie de estándares y que cada uno contenga el mismo nivel (nominal) de analito, y haciendo una serie de mediciones repetidas sobre cada estándar. Se puede utilizar ANOVA para verificar que la varianza entre estándares no sea significativa comparada con el error aleatorio de las mediciones y también para estimar esto último. Aquí “repetidas” significa una medición repetida de tal forma que se repiten *todas* las etapas del análisis, estimando de esta manera *todas* las posibles contribuciones al error aleatorio.

## BIBLIOGRAFÍA

Davies, O. L. y Goldsmith, P. L., *Statistical Methods in Research and Production*, Londres, Longmans, 1972. Proporciona ejemplos detallados de cómo se establece la capacidad de procesos en la construcción de diagramas de control.

Kateman, G. y Pijpers, F. E., *Quality Control in Analytical Chemistry*, Nueva York, Wiley, 1981. Proporciona un tratamiento muy detallado del muestreo.

Montgomery, D. C., *Introduction to Statistical Quality Control*, Nueva York, Wiley, 1985. Da un tratamiento detallado de todos los aspectos del control de la calidad.

Neave, H. R., *Elementary Statistics Tables*, Londres, Allen & Unwin, 1981. Contiene una tabla con una descripción detallada de sus usos para construir diagramas de Shewhart para  $\bar{x}$  y  $w$ .

Wernimont, G. T., *Use of Statistics to Develop and Evaluate Analytical Methods*, Spendley, W. (Ed.), Association of Official Analytical Chemists, Virginia, Arlington, 1985. Diseñado como el sucesor actualizado de Youden y Steiner (véase más abajo).

Woodward, R. H. y Goldsmith, P. L., *Cummulative Sum Techniques*, Monografía I.C.I. No. 3, Edinburgo, Oliver & Boyd, 1964. Un tratamiento detallado de DSA.

Youden, W. J. y Steiner, E. H., *Statistical Manual of the Association of Official Analytical Chemists*, Association of Official Analytical Chemists, Virginia, Arlington, 1975. Un manual clásico sencillo y bien escrito sobre los ensayos de colaboración.

Hartley, T. F., *Computerized Quality Control: Programs for the Analytical Laboratory*. Chichester, Horwood, 1987. Proporciona programas de computador para utilizarlos en control de calidad: los programas también se encuentran disponibles en disquetes.

## EJERCICIOS

1. Se proponen dos esquemas de muestreo para una situación en la que se sabe, por anteriores experiencias, que la varianza muestral es 10 y la varianza de medición es 4 (en unidades arbitrarias).

Esquema 1: Tome 5 incrementos de muestra, mézclelos y realice un análisis por duplicado.  
Esquema 2: Tome 3 incrementos de muestra y realice un análisis por duplicado de cada una.

Demuestre que la varianza de la medición es la misma para ambos esquemas.

¿Qué relación entre costo de muestreo y costo de análisis se debe sobrepasar para que el segundo esquema sea el más económico?

2. Los datos de la tabla adjunta dan la concentración de albúmina medida en el suero sanguíneo de un adulto. Se tomó una muestra de sangre durante cuatro días consecutivos y se determinó la concentración de albúmina en el suero por triplicado.

Día	Concentración de albúmina (normalizada, unidades arbitrarias)		
1	63	61	62
2	57	56	56
3	50	46	46
4	57	54	59

Demuestre que la concentración media para los diferentes días difiere significativamente. Estime la varianza de la variación día a día (es decir, “la variación muestral”).

3. Para estimar las varianzas de medición y muestrales se determinó la concentración de halofuginona en hígado de pollo, se tomaron cuatro incrementos de muestra de diferentes partes del hígado y se hicieron mediciones por triplicado sobre cada una de ellas. Se obtuvieron los siguientes resultados (mg/kg):

Muestra	Mediciones repetidas		
A	0.25	0.22	0.23
B	0.22	0.20	0.19
C	0.19	0.21	0.20
D	0.24	0.22	0.22

Verifique que la varianza muestral es significativamente más grande que la varianza de medición y estime ambas varianzas.

Se proponen dos esquemas disponibles de muestreo:

Esquema 1: Tome 6 incrementos de muestra, mézclelos y haga mediciones por cuadruplicado.  
Esquema 2: Tome 3 incrementos de muestra y sobre cada uno haga mediciones por duplicado.

Calcule la varianza total de la media para cada esquema.

4. Para estimar la capacidad de un proceso, se hicieron mediciones sobre 6 muestras de tamaño 4, como se muestra en la tabla adjunta (en la práctica se necesitarían al menos 25 de tales muestras). Estime la capacidad del proceso,  $\sigma$ . Si el valor objetivo es 50, calcule las posiciones de las líneas de aviso y de acción de un diagrama de Shewhart para la media muestral.

Muestras	Valores			
1	48.8	50.8	51.3	47.9
2	48.6	50.6	49.3	49.7
3	48.2	51.0	49.3	50.3
4	54.8	54.6	50.7	53.9
5	49.6	54.2	48.3	50.5
6	54.8	54.8	52.3	52.5

5. En un ensayo de colaboración, se enviaron dos muestras muy similares de aceite (A y B) a 15 laboratorios, cada uno de los cuales realiza una determinación del nivel de cadmio en cada muestra utilizando la técnica de emisión de plasma por acoplamiento inductivo. Se obtuvieron los siguientes resultados:

Laboratorio	Niveles de Cadmio (ppm)	
	A	B
1	8.8	10.0
2	3.8	4.7
3	10.1	12.1
4	8.0	11.0
5	5.0	4.7
6	5.2	6.4
7	6.7	8.7
8	9.3	9.6
9	6.9	7.5
10	3.2	2.8
11	9.7	10.4
12	7.2	8.3
13	6.5	6.8
14	9.7	7.2
15	5.0	6.0

Dibuje el diagrama de estas dos muestras para estos datos y comente la principal fuente de error en el ensayo de colaboración. Estime la varianza global, la varianza de medición, y el componente de error sistemático de la varianza de los resultados.

## Errores en análisis instrumental; regresión y correlación

### 5.1 ANÁLISIS INSTRUMENTAL

Las técnicas de análisis clásico o “química húmeda” como volumetrías y gravimetrías se utilizan todavía en muchos laboratorios y también se enseñan ampliamente en cursos de Química Analítica. Dichas técnicas suministran excelentes introducciones a la manipulación y otras prácticas requeridas en el trabajo analítico; son ideales para análisis de alta precisión, especialmente cuando se utiliza un pequeño número de muestras, y son necesarias para el análisis de materiales estándar. Sin embargo, en este momento no hay duda de que muchos análisis se efectúan por métodos instrumentales. Las técnicas que utilizan espectrometría de absorción y emisión a varias longitudes de onda, diferentes métodos electroquímicos, espectrometría de masas, cromatografía gaseosa y líquida, y métodos radioquímicos y térmicos, probablemente suponen el 90% de todo el trabajo analítico actual, o incluso más. Hay varias razones para esto.

(1) Los métodos instrumentales pueden realizar análisis que son difíciles o imposibles por los métodos clásicos. Mientras que éstos pueden detectar sólo rara vez especies químicas a niveles de submicrogramos, muchos métodos instrumentales son sorprendentemente sensibles. Por ejemplo, los métodos luminiscentes han detectado moléculas orgánicas a nivel de  $10^{-18}M$ . Por lo general, sólo es posible determinar un analito a la vez por métodos de “química húmeda”, si bien la espectrometría de plasma puede determinar diez o más elementos simultáneamente (y a muy bajas concentraciones), mientras que los métodos que combinan HPLC (cromatografía líquida de alta resolución, *high-performance liquid chromatography*) con detección espectroscópica pueden identificar y determinar muchos componentes de mezclas orgánicas complejas en pocos minutos. Aún más, el intervalo de concentración de un método concreto de análisis clásico está normalmente limitado por consideraciones teórico-prácticas. Así, titulaciones con AEDT (EDTA, en inglés) pueden realizarse con éxito con concentración de reactivo del orden de  $10^{-4}M$ , aunque se establece un límite superior ( $\sim 0.3M$ ) por la solubilidad del AEDT en agua. El

intervalo de concentración útil en métodos clásicos es generalmente de 2 a 3 órdenes de magnitud (es decir, potencias de 10). Por el contrario, algunos métodos instrumentales pueden determinar concentraciones de analito en el intervalo de seis o más órdenes de magnitud: esta característica tiene importantes repercusiones para el tratamiento estadístico de los resultados, como veremos en la próxima sección.

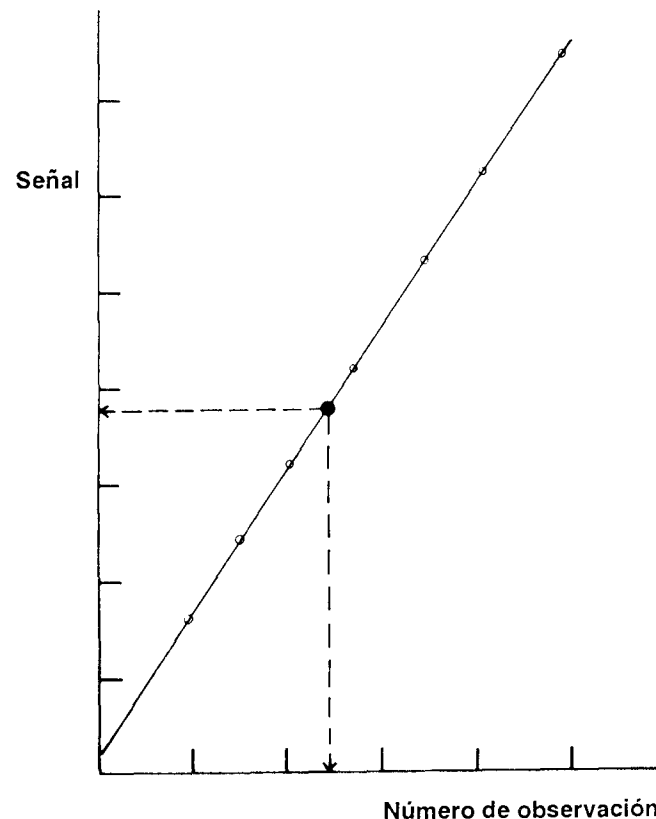
(2) Para una gran variedad de muestras, el análisis instrumental suele ser más rápido y a menudo más barato que los métodos manuales que requieren de mucha mano de obra. En análisis clínicos, por ejemplo, se presenta con frecuencia la necesidad de realizar el mismo análisis en una veintena o incluso en cientos de muestras de sangre o suero sanguíneo o plasma cada día; a pesar del alto costo inicial del equipo, esto, se realiza, por lo general, usando sistemas completamente automáticos. La automatización ha llegado a ser un aspecto tan importante en química analítica, que la facilidad con la cual puede automatizarse una técnica concreta a menudo determina si se utiliza o no.

Un método automático típico ofrece la posibilidad de procesar muestras a razón de cien por hora o más: el equipo tomará un volumen medido de muestra, la diluirá adecuadamente, provocará una o más reacciones, y determinará y registrará la concentración del analito o un derivado producido en las reacciones. Evidentemente en el análisis automático surgirán problemas especiales de estimaciones de error; los errores sistemáticos, por ejemplo, deben ser identificados y corregidos tan rápido como sea posible.

(3) Los instrumentos analíticos modernos pueden ponerse fácilmente en interfaz con computadores. La tendencia actual consiste en utilizar microcomputadores de 8 bits o 16 bits con dispositivos de almacenamiento de información de disco duro o blando. Tales sistemas pueden traer consigo un grado de perfeccionamiento razonable en el control del instrumento y manejo de datos. Por ejemplo, el computador puede controlar tiempos de muestreo o aparatos de inyección mediante activación eléctrica con conmutadores. Una vez que se han obtenido los resultados analíticos, el computador puede procesar los datos, realizando operaciones tales como la generación de derivadas de espectros o transformadas de Fourier. También puede evaluar los resultados estadísticos —el tema recurrente de este libro—, y comparar los resultados analíticos con datos almacenados en su memoria, así como comparar espectros y otros datos. Todos estos recursos se consiguen con computadores de bajo costo que funcionan a velocidades aceptables. Otra posibilidad importante es el desarrollo de instrumentos “inteligentes” que realizan, por ejemplo, procesos de optimización automáticos (véase el Cap. 7). No hay duda de que con los recursos de los computadores y su rápida incorporación a la gran mayoría de los instrumentos analíticos, éstos aumentarán su potencialidad y eficiencia.

## 5.2 GRÁFICAS DE CALIBRACIÓN EN ANÁLISIS INSTRUMENTAL

La capacidad de las técnicas de análisis instrumental para manejar un intervalo amplio de concentraciones de analito significa que se calculan los resultados, y se evalúan los errores aleatorios de una manera concreta, que difiere de la que se utiliza cuando se repite una sola medición varias veces. El procedimiento habitual es como sigue. El analista toma una serie de muestras (normalmente tres o cuatro, y posiblemente alguna más) en las que se *conoce* la concentración de analito. Estas calibraciones estándar se miden en el instrumento analítico en las mismas condiciones que las usadas para las muestras problema (es decir, las “desconocidas”). Una vez que se ha establecido la gráfica de calibración, se puede



**Figura 5.1** Procedimiento de calibración en análisis instrumental: ○ puntos de calibración; ● muestra problema.

obtener la concentración de analito en cualquier muestra problema por interpolación, como se indica en la figura 5.1. Este procedimiento general plantea varias preguntas estadísticas importantes.

- (1) ¿Es lineal la gráfica de calibración? Si es una curva, ¿qué forma tiene?
- (2) Teniendo en cuenta que cada uno de los puntos de la línea de calibración está sujeto a errores, ¿cuál es la mejor línea recta (o curva) que pasa por esos puntos?
- (3) Suponiendo que la calibración es realmente lineal, ¿cuáles son los errores estimados y los límites de confianza para la pendiente y la ordenada en el origen de la recta?
- (4) Cuando la gráfica de calibración se usa para el análisis de una muestra problema, ¿cuáles son los errores y los límites de confianza para la concentración determinada?
- (5) ¿Cuál es el *límite de detección* del método? Esto es, ¿cuál es la menor concentración de analito que se puede detectar con un nivel de confianza predeterminado?

Antes de abordar estas preguntas detalladamente, debemos considerar una serie de aspectos sobre el trazo de gráficas de calibración. Primero, por lo regular, es esencial que los estándares de calibración cubran el intervalo completo de concentraciones requerido en los análisis subsiguientes. Con la importante excepción del “método de las adiciones

estándar”, que se tratará en una sección posterior, la concentración de muestras problema se determina normalmente por interpolación y *no* por extrapolación. En segundo lugar, es muy importante incluir el valor de una muestra “en blanco” en la curva de calibración. La muestra en blanco no contiene ningún analito agregado deliberadamente, pero contiene los mismos disolventes, reactivos, etc., que las otras muestras problema, y está sujeta exactamente a la misma secuencia del procedimiento analítico. La señal del instrumento dada por la muestra en blanco no será a menudo cero. Desde luego, está sujeta a errores, como los otros puntos de la calibración, y no tiene sentido, en principio, sustraer el valor del blanco de los otros valores estándar antes de dibujar la gráfica de calibración. Finalmente, se debe subrayar que la curva de calibración se representa siempre con la respuesta del instrumento en el eje vertical ( $y$ ) y la concentración estándar sobre el eje horizontal ( $x$ ). Esto se debe a que muchos de los procedimientos que se describen en las secciones siguientes *suponen* que todos los errores se presentan en los valores de  $y$  y que las concentraciones estándares (valores de  $x$ ) se encuentran libres de error. Este supuesto se analiza posteriormente.

El lector debería considerar que los métodos descritos ahora siguen generando controversia, lo cual es patente en muchas reuniones científicas y trabajos originales acerca de la calibración. Muchas de estas controversias derivan del hecho de que los métodos utilizados generalmente se apoyan en dos supuestos. El primero es el que mencionamos en el párrafo anterior, o sea, que los errores en la calibración sólo ocurren en los valores de  $y$ . En muchos análisis instrumentales de rutina, este supuesto puede estar bien justificado. Los estándares pueden prepararse con un error de ca. 0.1% o inferior (véase el Cap. 1), mientras que las mediciones instrumentales podrían tener un coeficiente de variación del 1 al 2% o mayor. No obstante, con la aparición en los últimos años de algunos instrumentos automáticos de alta precisión nuevamente se han cuestionado estos supuestos, lo que ha llevado a algunos usuarios a preparar sus disoluciones estándar por peso en lugar de usar la combinación de peso y volumen, que es menos exacta.

El segundo supuesto es que la magnitud de los errores en  $y$  sea independiente de la concentración de analito. El sentido común indica que es improbable que esto sea verdad: si, por ejemplo, los errores relativos en la medición son constantes, los errores absolutos aumentarán a medida que aumenta la concentración de analito. Como veremos, es posible modificar el procedimiento estadístico teniendo en cuenta tales factores; sin embargo, en la práctica los procedimientos modificados (y bastante más complejos) parece que no se adoptan en muchos de los casos donde deberían utilizarse.

### 5.3 EL COEFICIENTE DE CORRELACIÓN MOMENTO-PRODUCTO

Aquí se analiza el primer problema planteado en la sección anterior: ¿es lineal la representación gráfica de la calibración? Supongamos que la representación de una línea recta toma la expresión algebraica:

$$y = bx + a \quad (5.1)$$

donde  $b$  es la pendiente de la recta y  $a$  su ordenada en el origen (punto donde corta al eje  $y$ ). Los puntos individuales sobre la línea se denotarán por  $(x_1, y_1)$  (normalmente la lectura del “blanco”),  $(x_2, y_2)$ ,  $(x_3, y_3) \dots (x_i, y_i) \dots (x_n, y_n)$ , es decir, hay  $n$  puntos. La media de los

valores de  $x$ , por lo general, se denomina  $\bar{x}$  y la media de los valores de  $y$ ,  $\bar{y}$ : la posición  $(\bar{x}, \bar{y})$  se conoce como el “centro de gravedad” de todos los puntos.

Para estimar si los puntos experimentales se ajustan bien o no a una línea recta, calculamos el **coeficiente de correlación momento-producto**,  $r$ . Para simplificar, denominaremos a este dato estadístico como “coeficiente de correlación” debido a que en las ciencias cuantitativas es el más usado; sin embargo, encontraremos otros coeficientes de correlación en el capítulo 6. El valor de  $r$  está dado por:

$$r = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\{[\sum_i (x_i - \bar{x})^2][\sum_i (y_i - \bar{y})^2]\}^{1/2}} \quad (5.2)$$

Un minucioso estudio de esta ecuación muestra que  $r$  puede tomar valores en el intervalo de  $-1 \leq r \leq +1$ . Como se indica en la figura 5.2, un valor de  $r$  de  $-1$  describe una correlación negativa perfecta, es decir, todos los puntos experimentales están en la línea recta de pendiente negativa. En forma similar, cuando  $r = +1$ , tenemos una correlación positiva perfecta, ya que todos los puntos están exactamente en una línea recta de pendiente

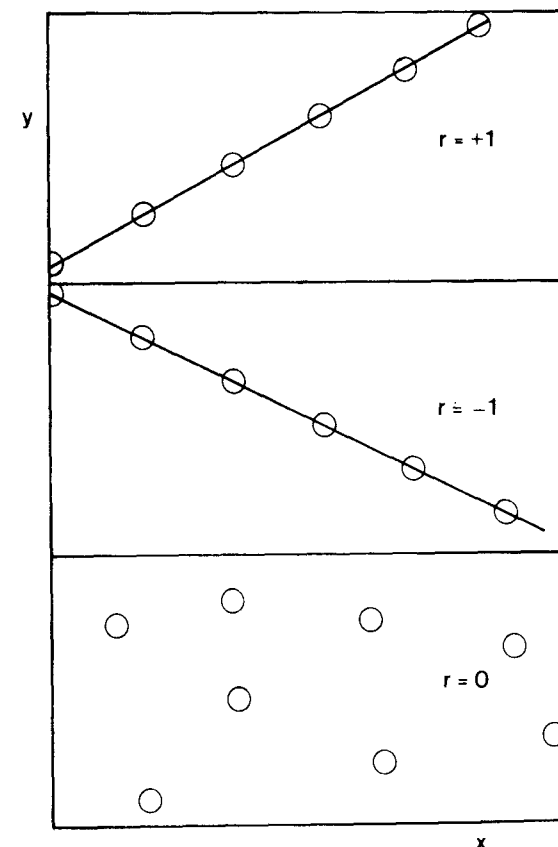


Figura 5.2 El coeficiente de correlación momento-producto,  $r$ .

positiva. Cuando no existe correlación entre  $x$  y  $y$ , el valor de  $r$  es cero. En la práctica analítica, las gráficas de calibración proporcionan frecuentemente valores numéricos de  $r$  mayores que 0.99; los valores de  $r$  menores que 0.90 son relativamente poco comunes. Un ejemplo típico de un cálculo de  $r$  ilustra una serie de aspectos importantes.

*Ejemplo.* Se ha examinado una serie de soluciones estándar de fluoresceína en un fluorímetro, y condujo a las siguientes intensidades de fluorescencia (en unidades arbitrarias):

Intensidad de fluorescencia:	2.1	5.0	9.0	12.6	17.3	21.0	24.7
Concentración, pg/ml:	0	2	4	6	8	10	12

Determine el coeficiente de correlación,  $r$ .

En la práctica, estos cálculos seguramente se realizarán en una calculadora programable o en un microcomputador, pero es importante, e instructivo, examinar un resultado calculado manualmente. Los datos se presentan en una tabla, como la que sigue.

$x_i$	$y_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
0	2.1	-6	36	-11.0	121.00	66.0
2	5.0	-4	16	-8.1	65.61	32.4
4	9.0	-2	4	-4.1	16.81	8.2
6	12.6	0	0	-0.5	0.25	0
8	17.3	2	4	4.2	17.64	8.4
10	21.0	4	16	7.9	62.41	31.6
12	24.7	6	36	11.6	134.56	69.6
42	91.7	0	112	0	418.28	216.2

$$\bar{x} = 42/7 = 6; \bar{y} = 91.7/7 = 13.1$$

Los números que se encuentran por debajo de la línea, al pie de las columnas, son en cada caso las sumas de los números de la tabla. [Note que  $\sum (x_i - \bar{x})$  y  $\sum (y_i - \bar{y})$  ambos son cero.] Usando estos totales y la ecuación (5.2), tenemos

$$\begin{aligned} r &= 216.2 / (112 \times 418.28)^{1/2} \\ &= 216.2 / 216.44 \\ &= 0.9989 \end{aligned}$$

Dos observaciones se siguen de este ejemplo. Como se muestra en la figura 5.3, aunque varios de los puntos se desvían notablemente de la "mejor" línea recta (calculada utilizando los principios de la sección siguiente), el valor de  $r$  es muy cercano a 1. La experiencia muestra que aun con gráficas de calibración bastante pobres a la vista, los valores correspondientes de  $r$  son altos. En tales casos, el numerador y el denominador en la ecuación (5.2) son aproximadamente iguales. Entonces es muy importante realizar el cálculo con un número adecuado de cifras significativas. En el ejemplo dado, despreciar las cifras después del punto decimal daría un valor incorrecto de  $r$  de exactamente 1, y utilizar sólo una cifra decimal supondría un valor incorrecto de  $r$  de 0.9991. Este aspecto es especialmente importante cuando se usa una calculadora o microcomputador para determinar  $r$ ; tales herramientas no siempre proporcionan suficientes cifras.

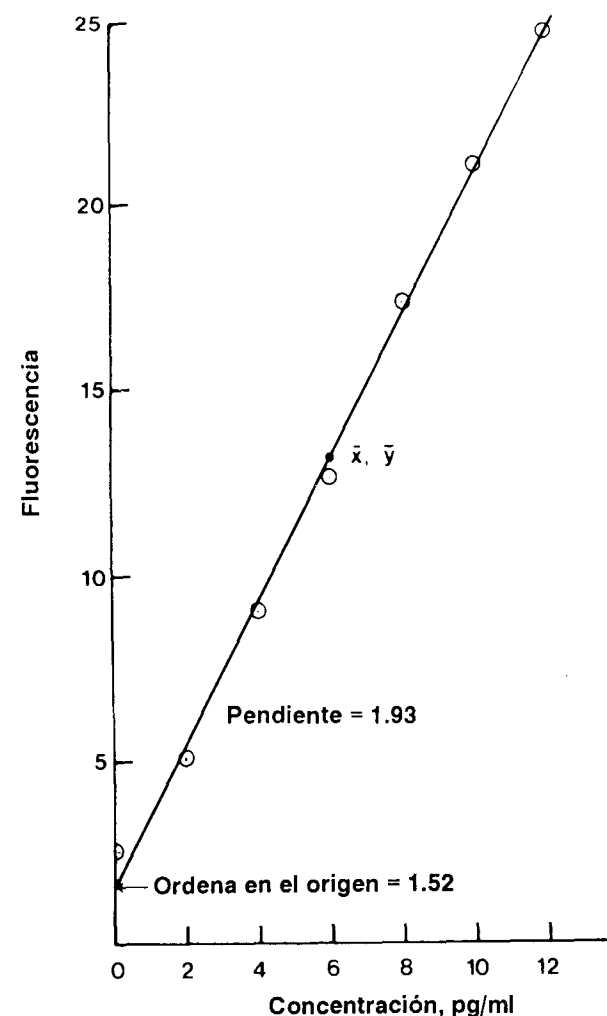


Figura 5.3 Gráfica de calibración de los datos dE LA PÁg. 92.

Aunque los coeficientes de correlación son sencillos de calcular, también son mal interpretados con facilidad. Debe tenerse en cuenta que el uso de la ecuación (5.2) generará un valor de  $r$  aun cuando los datos sean claramente de carácter no lineal. En la figura 5.4 se muestran dos ejemplos en los que un cálculo de  $r$  llevaría a conclusiones erróneas. En la figura 5.4(a), los puntos de la gráfica de calibración pertenecen a una curva; esta curva es suficientemente suave, aunque se obtiene un alto coeficiente de correlación cuando se aplica la ecuación (5.2). De este ejemplo deducimos que *siempre* hay que representar físicamente la curva de calibración (en papel milimetrado o en monitor de computador): si no es así, del cálculo de  $r$  se puede deducir erróneamente una relación de carácter lineal. La figura 5.4(b) es una advertencia de que un coeficiente de correlación cero no significa exactamente que  $y$  y  $x$  no estén relacionados, sino que no están *linealmente* relacionados.

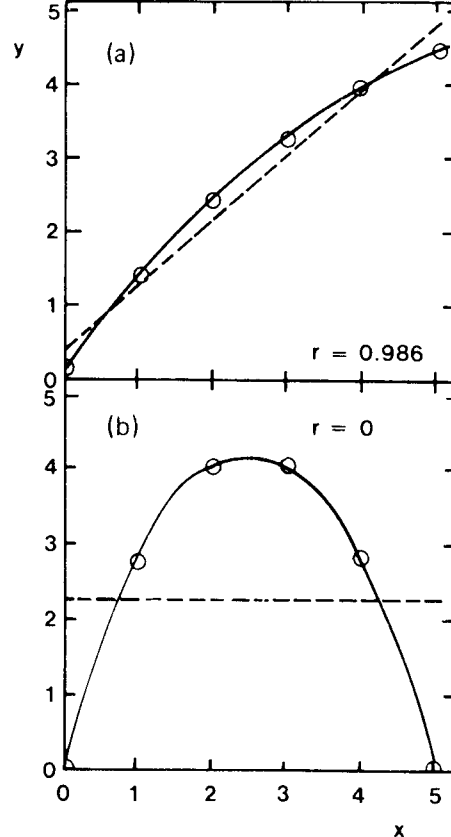


Figura 5.4 Interpretaciones erróneas del coeficiente de correlación,  $r$ .

Como hemos visto, los valores de  $r$  obtenidos en análisis instrumental son generalmente muy altos, de manera que un valor calculado, junto con la propia gráfica de calibración, suele ser suficiente para asegurar al analista que de hecho ha obtenido una relación lineal útil. Sin embargo, en algunas circunstancias, se obtiene valores de  $r$  mucho más bajos; una de estas situaciones se analizará posteriormente en la sección 5.9. En estos casos será necesario utilizar una prueba estadística adecuada para ver si el coeficiente de correlación es realmente significativo, considerando el número de pares de valores usados en su cálculo. El método más simple es calcular un valor  $t$  (véase el capítulo 3 para un estudio más completo de la prueba  $t$ ), usando la ecuación

$$t = \frac{|r| \sqrt{(n-2)}}{\sqrt{(1-r^2)}} \quad (5.3)$$

El valor calculado de  $t$  se compara con el tabulado al nivel de significación deseado, usando una prueba  $t$  de *dos colas* y  $(n-2)$  grados de libertad. La hipótesis nula en este caso es que no existe correlación entre  $x$  y  $y$ . Si el valor calculado de  $t$  es mayor que el tabulado, la hipótesis nula se rechaza; esto es, se concluye que existe una correlación significativa.

## 5.4 LA RECTA DE REGRESIÓN DE $y$ SOBRE $x$

En esta sección, suponemos que existe una relación lineal entre la señal analítica ( $y$ ) y la concentración ( $x$ ), y mostramos cómo calcular la “mejor” línea recta a través de los puntos de la gráfica de calibración, cada uno de los cuales está sujeto a un error experimental. Hasta aquí hemos supuesto que todos los errores se encuentran en  $y$  (véase la Sec. 5.2), ahora trataremos de buscar la recta que minimice las desviaciones en la dirección  $y$  entre los puntos experimentales y la línea calculada. Ya que algunas de esas desviaciones (técnicamente conocidas como los residuos de  $y$ ) serán positivas y otras negativas, es razonable intentar minimizar la **suma de los cuadrados de los residuos**. Esto explica el uso frecuente del término “método de los mínimos cuadrados” para este procedimiento. La línea recta buscada se calcula basándose en este principio: la línea debe pasar por el “centro de gravedad” de los puntos,  $(\bar{x}, \bar{y})$ . Puede demostrarse que

$$b = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2} \quad (5.4)$$

$$a = \bar{y} - b\bar{x} \quad (5.5)$$

La línea calculada de esta forma se conoce como la **recta de regresión de  $y$  sobre  $x$** , es decir, la recta indica cómo varía  $y$  cuando  $x$  se ajusta a los valores elegidos. Es muy importante notar que la recta de regresión de  $x$  sobre  $y$  *no es la misma recta* (excepto en el caso muy improbable donde todos los puntos caigan sobre una línea recta, exactamente cuando  $r = 1$ ). La recta de regresión de  $x$  sobre  $y$  (que también pasa por el centro de gravedad de todos los puntos) supone que todos los errores ocurren en la dirección  $x$ . Si se mantiene rigurosamente el convenio de que la señal analítica se representa siempre sobre el eje  $y$  y la concentración sobre el eje  $x$ , debemos usar siempre la recta de regresión de  $y$  sobre  $x$  en los experimentos de calibración.

*Ejemplo.* Calcule la pendiente y ordenada en el origen de la recta de regresión para los datos expuestos en el ejemplo anterior (Sec. 5.3).

En la sección 5.3 calculamos que, para esta curva de calibración:

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) = 216.2; \quad \sum_i (x_i - \bar{x})^2 = 112; \quad \bar{x} = 6; \quad \bar{y} = 13.1.$$

Usando la ecuaciones (5.4) y (5.5) calculamos que

$$b = 216.2/112 = 1.93$$

$$a = 13.1 - (1.93 \cdot 6) = 13.1 - 11.58 = 1.52$$

La ecuación para la recta de regresión es  $y = 1.93x + 1.52$ .

Es obvio que los términos utilizados para calcular  $r$  son necesarios también en el cálculo de  $a$  y  $b$ : esto se facilita con una calculadora o un programa de computador sencillo.

Los resultados de los cálculos de la pendiente y ordenada en el origen se representa en la figura 5.3. De nuevo es importante recalcar que la ecuaciones (5.4) y (5.5) no deben



emplearse mal porque sólo darán resultados útiles cuando un estudio previo (cálculo de  $r$  y una inspección visual de los puntos) indique que una línea recta es coherente con el experimento que se trata. Para calcular rectas de regresión pueden usarse también métodos no paramétricos (es decir, métodos que no hacen suposiciones acerca de la naturaleza de la distribución del error): este tema se trata en el próximo capítulo.

## 5.5 ERRORES EN LA PENDIENTE Y ORDENADA EN EL ORIGEN DE LA RECTA DE REGRESIÓN

La recta de regresión calculada en la acción anterior se utilizará en la práctica para estimar la concentración de las muestras problema por interpolación, y quizás también para estimar el límite de detección del procedimiento analítico. Los errores aleatorios en los valores de la pendiente y de la ordenada en el origen son importantes, por lo que consideramos ahora las ecuaciones utilizadas para calcularlos. Primero debemos calcular el dato estadístico  $s_{y/x}$ , que está dado por:

$$s_{y/x} = \left\{ \frac{\sum_i (y_i - \hat{y}_i)^2}{n - 2} \right\}^{\frac{1}{2}} \quad (5.6)$$

Se comprueba que esta ecuación utiliza los residuos de  $y$ ,  $y_i - \hat{y}_i$ , donde los valores  $\hat{y}_i$  son los puntos sobre la recta de regresión calculada, correspondientes a los valores individuales de  $x$  es decir, los valores de  $y$  “ajustados” (Fig. 5.5). El valor de  $\hat{y}$  para un valor de  $x$  dado se calcula rápidamente a partir de la ecuación de regresión. La forma de la ecuación (5.6) es similar a la de la ecuación de la desviación estándar de una serie de mediciones repetidas [Ec. (2.2)]; pero difiere en que las desviaciones  $(y_i - \bar{y})$ , son sustituidas por los residuos,  $(y_i - \hat{y}_i)$ , y el denominador contiene el término  $(n - 2)$  en vez de  $(n - 1)$ . En un cálculo de regresión lineal el número de grados de libertad (véase la Sec. 2.4) es  $(n - 2)$ . Esto refleja el hecho obvio de que para representar una línea recta sólo se necesitan dos puntos.

Después de obtener un valor de  $s_{y/x}$ , podemos calcular  $s_b$  y  $s_a$ , las desviaciones estándar para la pendiente ( $b$ ) y la ordenada en el origen ( $a$ ). Éstas están dadas por:

$$s_b = \frac{s_{y/x}}{\left\{ \sum_i (x_i - \bar{x})^2 \right\}^{\frac{1}{2}}} \quad (5.7)$$

$$s_a = s_{y/x} \left\{ \frac{\sum_i x_i^2}{n \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \quad (5.8)$$

Los valores de  $s_b$  y  $s_a$  se pueden utilizar de la manera usual (véase el Cap. 2) para estimar límites de confianza para la pendiente y la ordenada en el origen. Así, los límites de confianza para la pendiente están dados por  $b \pm t s_b$ , donde el valor de  $t$  se obtiene para un nivel de confianza deseado y  $(n - 2)$  grados de libertad. De manera similar, los límites de confianza para la ordenada en el origen están dados por  $a \pm t s_a$ .

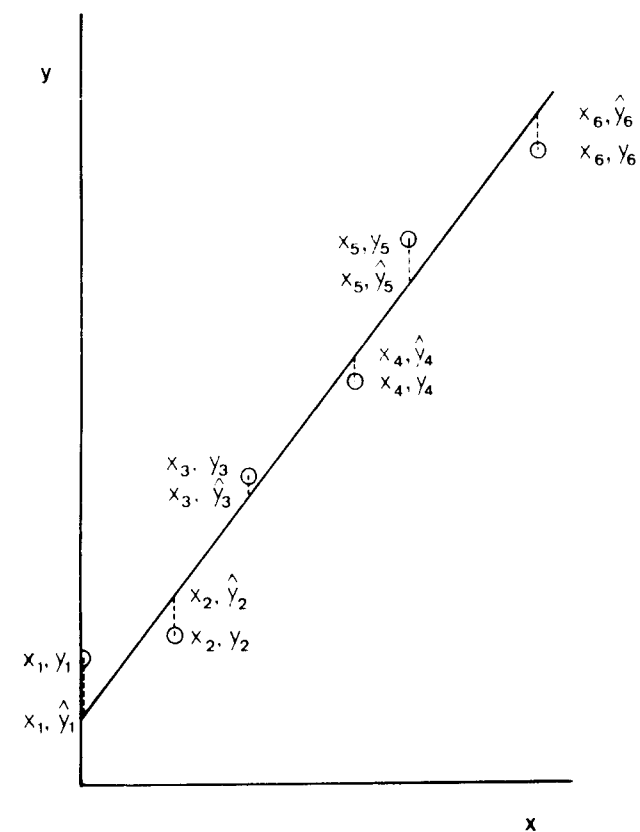


Figura 5.5 Los residuos de  $y$  de una regresión lineal.

*Ejemplo.* Calcule la desviación estándar y los límites de confianza para la pendiente y la ordenada en el origen de la recta de regresión calculada en la sección 5.4.

Es posible que no realicemos este tipo de cálculo directamente en una calculadora programable, aunque se han desarrollado programas de computador útiles (véase la bibliografía del capítulo 1). Aquí se efectúa el cálculo manualmente, utilizando una tabla estructurada como antes.

$x_i$	$x_i^2$	$y_i$	$\hat{y}_i$	$ y_i - \hat{y}_i $	$(y_i - \hat{y}_i)^2$
0	0	2.1	1.52	0.58	0.3364
2	4	5.0	5.38	0.38	0.1444
4	16	9.0	9.24	0.24	0.0576
6	36	12.6	13.10	0.50	0.2500
8	64	17.3	16.96	0.34	0.1156
10	100	21.0	20.82	0.18	0.0324
12	144	24.7	24.68	0.02	0.0004
$\sum_i x_i^2 = 364$				$\sum_i (y_i - \hat{y}_i)^2 = 0.9368$	

A partir de la tabla, y utilizando la ecuación (5.6) se obtiene

$$s_{y/x} = \sqrt{0.9368/5} = \sqrt{0.18736} = 0.4329.$$

De la sección 5.3 sabemos que  $\sum_i (x_i - \bar{x})^2 = 112$ , y la ecuación (5.7) puede utilizarse para mostrar que

$$s_b = 0.4329/\sqrt{112} = 0.4329/10.58 = 0.0409.$$

El valor de  $t$  para  $(n - 2) = 5$  y el nivel de confianza del 95% es 2.7 (tabla A.1). Entonces los límites de confianza al 95% para  $b$  son

$$b = 1.93 \pm 2.57 \times 0.0409 = 1.93 \pm 0.11$$

La ecuación (5.8) requiere conocer  $\sum x_i^2$ , calculado de la tabla como 364. Así, podemos escribir

$$s_a = 0.4329\sqrt{(364/784)} = 0.2950$$

de manera que los límites de confianza son

$$a = 1.52 \pm 2.57 \times 0.2950 = 1.52 \pm 0.76$$

En este ejemplo, el número necesario de cifras significativas no es grande, pero es conveniente utilizar siempre el número máximo disponible durante el cálculo, redondeando sólo al final.

## 5.6 CÁLCULO DE UNA CONCENTRACIÓN

Una vez determinadas la pendiente y la ordenada en el origen de la recta de regresión es fácil calcular un valor de  $x$  correspondiente a cualquier valor de  $y$  medido. Pero surge un problema más complejo cuando necesitamos estimar el error en una concentración calculada utilizando una recta de regresión. El cálculo de un valor de  $x$  para un valor de  $y$  dado conlleva el uso de la pendiente ( $b$ ) y la ordenada en el origen ( $a$ ) y, como vimos en la sección anterior, ambos valores están sujetos a error. Como resultado, la determinación del error en el valor de  $x$  es extremadamente compleja y muchos profesionales utilizan la siguiente fórmula aproximada:

$$s_{x_0} = \frac{s_{y/x}}{b} \left\{ 1 + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \quad (5.9)$$

En esta ecuación  $y_0$  es el valor experimental de  $y$  a partir del cual se determina el valor  $x_0$  de la concentración;  $s_{x_0}$  es la desviación estándar de  $x_0$ , y los otros símbolos tienen su significado usual. En algunos casos, un analista puede realizar varias lecturas para obtener el valor de  $y_0$ ; si se dispone de  $m$  lecturas, entonces la ecuación para  $s_{x_0}$  se convierte en

$$s_{x_0} = \frac{s_{y/x}}{b} \left\{ \frac{1}{m} + \frac{1}{n} + \frac{(y_0 - \bar{y})^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \quad (5.10)$$

Como se esperaba, la ecuación (5.10) se reduce a la (5.9) si  $m = 1$ . Como siempre, los límites de confianza pueden calcularse como  $x_0 \pm t s_{x_0}$ , con  $(n - 2)$  grados de libertad. Como mencionamos antes, un programa de computador simple realizará todos estos cálculos, aunque muchas calculadoras preprogramables no serán del todo adecuadas.

*Ejemplo.* Usando los datos del ejemplo de la sección 5.3, determine los valores de  $x_0$  y  $s_{x_0}$  y los límites de confianza de  $x_0$  para soluciones con intensidades de fluorescencia de 2.9, 13.5 y 23.0 unidades.

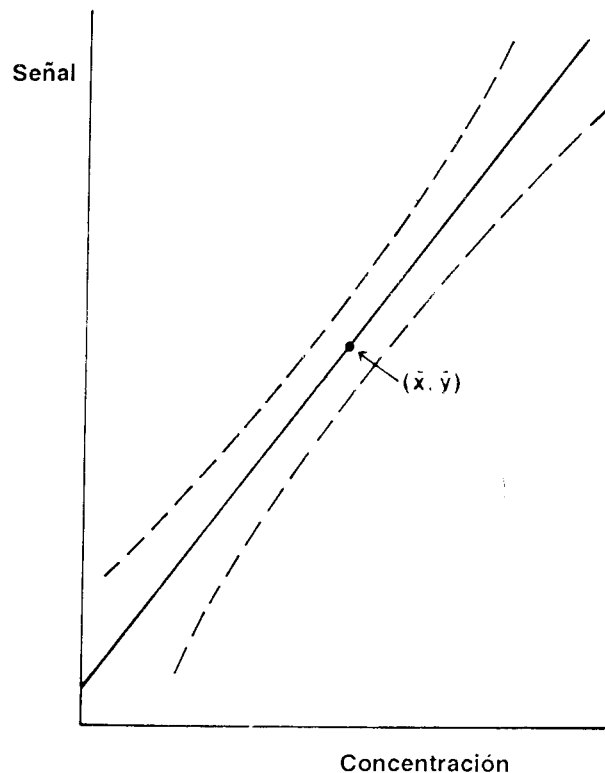
Los valores de  $x_0$  se calculan fácilmente utilizando la ecuación de regresión determinada en la sección 5.4,  $y = 1.93x + 1.52$ . Sustituyendo los valores de  $y_0$  2.9, 13.5 y 23, obtenemos los valores de  $x_0$  de 0.72, 6.21 y 11.3 pg/ml respectivamente.

Para obtener los valores de  $s_{x_0}$  correspondientes a los valores de  $x_0$  utilizamos la ecuación (5.9), recordando de las secciones precedentes que  $n = 7$ ,  $b = 1.93$ ,  $s_{y/x} = 0.4329$ ,  $\bar{y} = 13.1$ , y  $\sum_i (x_i - \bar{x})^2 = 112$ . Los valores de  $y_0$  2.9, 13.5 y 23.0 proporcionan unos valores de  $s_{x_0}$  de 0.26, 0.24 y 0.26, respectivamente. Los límites de confianza correspondientes al nivel del 95% ( $t = 2.57$ ) son  $0.72 \pm 0.68$ ,  $6.21 \pm 0.62$  y  $11.3 \pm 0.68$  pg/ml respectivamente.

Este ejemplo aclara un aspecto de cierta importancia. Es obvio que los límites de confianza son bastante más pequeños (es decir mejores) para el resultado de  $y_0 = 13.5$  que para los otros dos valores de  $y_0$ . El estudio de la ecuación (5.9) confirma que mientras que  $y_0$  se aproxima a  $\bar{y}$ , el tercer término dentro del paréntesis se aproxima a cero, y  $y_0$  al valor mínimo. La forma general de los límites de confianza para una concentración calculada se muestra en la figura 5.6. Además, en un análisis práctico un experimento de calibración de este tipo proporcionará los resultados más precisos cuando la señal del instrumento medida corresponda a un punto próximo al centro de gravedad de la recta de regresión.

Si deseamos mejorar (es decir reducir) los límites de confianza en este experimento de calibración, las ecuaciones (5.9) y (5.10) muestran que se puede considerar al menos dos aproximaciones. Podemos aumentar  $n$ , el número de puntos de la calibración de la recta de regresión, y podemos hacer más de una medición de  $y_0$ , y utilizar el valor medio de dichas  $m$  mediciones en el cálculo de  $x_0$ . Los resultados de tales procedimientos pueden establecerse considerando los tres términos dentro de los paréntesis, en las dos ecuaciones. En el ejemplo anterior, el término predominante en los tres cálculos es el primero, o sea la unidad. Se sigue que en este caso (y en muchos otros) puede llevarse a cabo una mejoría en la precisión midiendo  $y_0$  varias veces y usando la ecuación (5.10) en lugar de la ecuación (5.9). Si, por ejemplo, se ha calculado el valor de  $y_0$  de 13.5 como la media de cuatro determinaciones, entonces el valor de  $s_{x_0}$  y los límites de confianza serían 0.14 y  $6.21 \pm 0.36$  respectivamente; ambos resultados indican que se ha mejorado sustancialmente la precisión. Por supuesto, si se hacen muchas mediciones repetidas (suponiendo que se dispone de suficiente muestra) se genera mucho más trabajo y se obtiene sólo un beneficio adicional pequeño: el lector puede verificar que 8 mediciones de  $y_0$  producirán un valor de  $s_{x_0}$  de 0.12 y unos límites de confianza de  $6.21 \pm 0.30$ .

El efecto de  $n$ , el número de puntos de calibración, sobre los límites de confianza en la determinación de la concentración es más complejo. Esto se debe también a que tenemos que considerar la presencia de cambios en el valor de  $t$ . El uso de un gran número de muestras de calibración supone preparar muchos estándares exactos para incrementar sólo marginalmente la precisión (véanse los efectos de incrementar  $m$ , descritos en el párrafo



**Figura 5.6** Forma general de los límites de confianza para una concentración determinada utilizando una recta de regresión no ponderada.

anterior). Por otra parte, no es adecuado usar valores de  $n$  pequeños: en tal caso, no sólo  $1/n$  será grande, sino que el número de grados de libertad  $n - 2$ , se hará muy pequeño, y se necesitarán valores muy grandes de  $t$  en el cálculo de los límites de confianza. En muchos experimentos, como en el ejemplo dado, aproximadamente seis puntos de calibración serán suficientes; si es necesario el analista ganará precisión extra repitiendo mediciones de  $y_0$ .

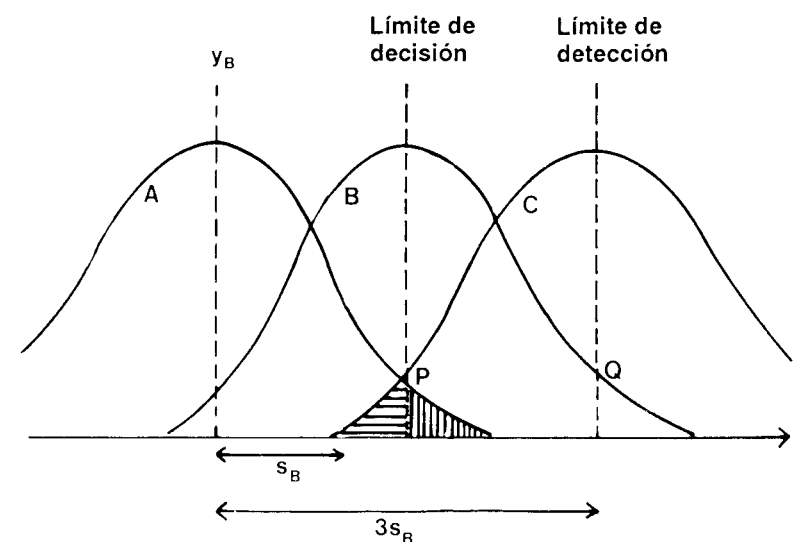
## 5.7 LÍMITES DE DETECCIÓN

Como hemos visto, una de las ventajas de utilizar métodos instrumentales de análisis es que son capaces de detectar y determinar cantidades de analito mucho más pequeñas que los métodos de análisis clásicos. Esta ventaja permite apreciar la importancia de las concentraciones en el nivel de trazas en muchos materiales, por ejemplo en muestras biológicas y medioambientales, contribuyendo así al desarrollo de muchas otras técnicas en las cuales los límites de detección bajos son el mejor criterio para aplicarlas con éxito. Es evidente que los métodos estadísticos son importantes para la evaluación y la comparación de límites de detección. En términos generales, se puede describir el límite de detección de un analito como aquella concentración que proporciona una señal en el instrumento ( $y$ ) *significativamente diferente* de la señal de una muestra en "blanco" o "señal

de fondo". Es obvio que esta descripción da al analista un buen margen de libertad para decidir la definición exacta del límite de detección, basada en una adecuada interpretación de la frase "significativamente diferente". En la práctica, existe muy poco acuerdo entre los profesionales y los organismos oficiales sobre este punto, por lo que esta área resulta tan controvertida como otros aspectos del tratamiento estadístico del análisis de calibración. Una definición que se utiliza comúnmente en la bibliografía de la química analítica es que el límite de detección es *la cantidad de concentración de analito que proporciona una señal igual a la señal del blanco,  $y_B$ , más dos veces la desviación estándar del blanco,  $s_B$* . Recientes recomendaciones de organismos públicos (particularmente estadounidenses) sugieren que el criterio debería de ser:

$$y - y_B = 3s_B \quad (5.11)$$

La significación de esta última definición se aclara con mayor detalle en la figura 5.7. Un analista que estudia concentraciones traza se enfrenta a dos problemas: por un lado no desea dar un informe de la presencia de analito cuando esté realmente ausente, pero por otro, tampoco pretende informar que el analito está ausente cuando de hecho está presente. La posibilidad de que ocurra cada uno de estos errores debe minimizarse bajo una definición lógica de un límite de detección. En la figura, la curva A representa la distribución normal de valores medidos de la señal del blanco. Sería posible identificar un punto  $y = P$ , hacia el borde superior de esta distribución, y pretender que sería improbable que una señal más grande que ésta se debiese al blanco (Fig. 5.7), mientras que se supondría que una señal menor que  $P$  indica una muestra en blanco. Sin embargo, para que una muestra dé una señal promedio  $P$ , el 50% de las señales observadas será menor que ésta, ya que tendrá una distribución normal (de la misma forma que para el blanco, véase más adelante) que se prolonga por debajo de  $P$  (curva B). La probabilidad de concluir que esta muestra no difiere de la del blanco cuando de hecho *lo hace* es por lo tanto del 50%. El punto  $P$ ,



**Figura 5.7** Definiciones de (a) el límite de decisión y (b) el límite de detección.

llamado el límite de decisión, es insatisfactorio como un límite de detección, ya que soluciona el primero de los problemas mencionados antes, pero no el segundo. Un punto más adecuado está en  $y = Q$  (véase la Fig. 5.7), de manera que  $Q$  está situado al doble de la distancia que existe entre  $y_B$  y  $P$ . Se puede demostrar que si  $y_B - Q$  es 3.28 veces la desviación estándar del blanco,  $s_B$ , entonces la probabilidad de que cada uno de los dos tipos de error ocurra (indicados por las áreas sombreadas de la figura 5.7) es sólo del 5%. Si, como se sugiere en la ecuación (5.11), la distancia  $y_B - Q$  es sólo  $3s_B$ , la probabilidad de cada error es aproximadamente del 7%; muchos analistas deberían considerar que esta definición del límite de detección es razonable.

Es preciso insistir de nuevo en que esta definición es completamente arbitraria, y está abierta para que cualquier analista proponga una alternativa ante un propósito concreto. Por ejemplo, existen ocasiones en que un analista está ansioso de evitar a toda costa la posibilidad de informar sobre la ausencia de analito cuando de hecho está presente, pero lo contrario no es relativamente preocupante. Está claro que cuando se cite un límite de detección en un trabajo o informe, debe proporcionarse la definición utilizada para obtenerlo. Se han realizado algunos intentos para definir otro límite, el "límite de cuantificación" (o "límite de determinación"), considerado como el límite más bajo para mediciones cuantitativamente precisas, en oposición a la detección cualitativa. Se ha sugerido un valor de  $y_B + 10s_B$  para este límite, aunque en la práctica no se ha utilizado ampliamente.

A continuación analizaremos cómo se obtienen en la práctica los términos  $y_B$  y  $s_B$  cuando se utiliza una recta de regresión convencional para la calibración, tal como se describió en las secciones precedentes. Un supuesto fundamental del método de los mínimos cuadrados sin ponderar que ya hemos estudiado en este capítulo es que cada punto en la representación gráfica (incluido el que representa al blanco o ruido de fondo) tiene una variación distribuida normalmente (sólo en la dirección de  $y$ ) con una desviación estándar estimada por  $s_{y/x}$  [véase la Ec. (5.6)]. Ésta es la justificación para representar las curvas de distribución normal con la misma amplitud que refleja la figura 5.7; por lo tanto es apropiado utilizar  $s_{y/x}$  en lugar de  $s_B$  en la estimación del límite de detección. Por supuesto, se puede realizar el experimento con el blanco varias veces y obtener un valor independiente para  $s_B$ , pero es un procedimiento que lleva tiempo, y en la práctica el uso de  $s_{y/x}$  es bastante adecuado. El valor de  $a$ , la ordenada en el origen calculada, puede utilizarse como una estimación de  $y_B$ , la propia señal del blanco; debería ser una estimación más exacta de  $y_B$  que el solo valor de la medición del blanco,  $y_1$ .

**Ejemplo.** Estime el límite de detección para determinar la fluoresceína estudiada en las secciones anteriores.

Se utiliza la ecuación (5.11) con los valores de  $y_B (= a)$  y  $s_B (= s_{y/x})$  calculados previamente. El valor de  $y$  en el límite de detección está dado por  $1.52 + 3 \times 0.4329$ , es decir, 2.82. El uso de la ecuación de regresión conduce a un límite de detección de 0.67 pg/ml. La figura 5.8 resume todos los cálculos realizados a partir de los datos para determinar la fluoresceína.

Es muy importante evitar la confusión entre el límite de detección de una técnica con su **sensibilidad**. Esta confusión surge, probablemente, porque no existe una palabra que signifique "tener un límite de detección bajo"; la palabra "sensible" se utiliza generalmente con este propósito, lo que da lugar a demasiada ambigüedad. La sensibilidad de una técnica se define correctamente como la *pendiente* de una línea de calibración y, siempre que la

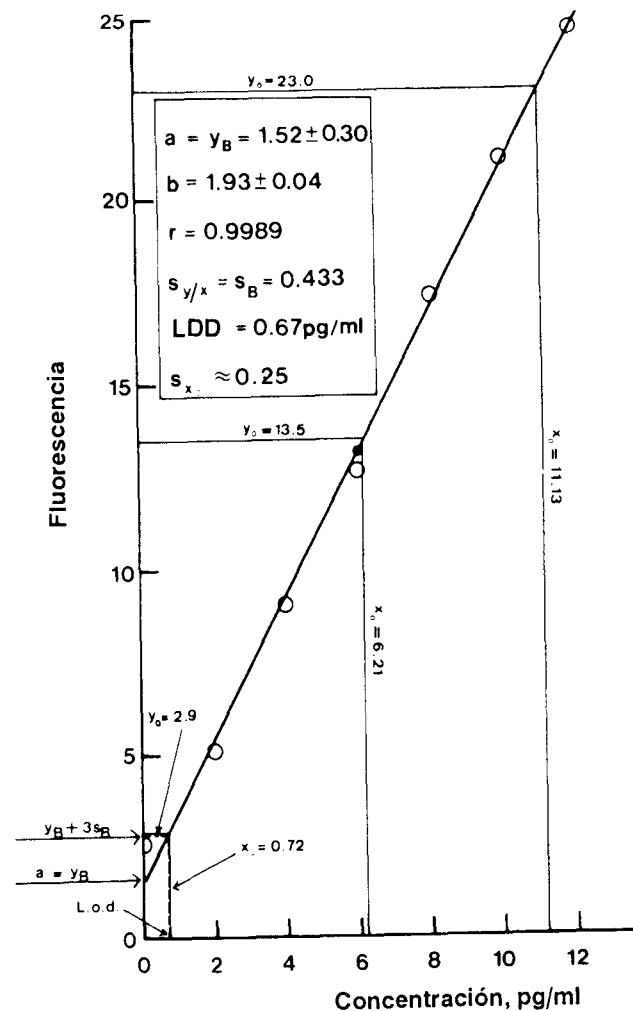


Figura 5.8 Resumen de los cálculos usando los datos de la pág. 92.

representación sea lineal, puede ser medida en cualquier punto de ella. Por el contrario, el límite de detección de un método se calcula con ayuda de la zona de la representación cercana al origen, y utiliza tanto la pendiente como la ordenada al origen.

## 5.8 EL MÉTODO DE LAS ADICIONES ESTÁNDAR

Supongamos que un analista desea determinar la concentración de plata en muestras de desechos fotográficos por espectrometría de absorción atómica. Si utiliza los métodos de las secciones anteriores, podría efectuar una calibración en el espectrofotómetro con soluciones acuosas de una sal de plata pura y utilizar la gráfica de calibración resultante en la determinación de plata en las muestras problema. Sin embargo, este método sólo es válido si una solución acuosa pura de plata da la misma señal de absorción que una muestra

fotográfica de desecho que contenga la misma concentración de plata. En otras palabras, cuando se usan soluciones puras para establecer la gráfica de calibración se supone que no hay “efectos de matriz”, es decir, disminución o aumento en la señal de absorbancia de plata debido a la presencia de otros componentes en la muestra. Frecuentemente, tal suposición no es válida en muchas áreas del análisis. Los efectos de matriz ocurren incluso con métodos tales como la espectrometría de plasma que tienen la ventaja de estar relativamente libres de interferencias.

La primera solución posible a este problema es tomar una muestra de desechos fotográficos que sea similar a la muestra problema pero libre de plata, y añadirle cantidades conocidas de una sal de plata para formar soluciones estándar. Entonces, la gráfica de calibración se establecerá utilizando una matriz aparentemente adecuada; sin embargo, en muchos casos esta aproximación es impracticable. No se eliminarán efectos de matriz que difieran en magnitud de una muestra a otra, y quizá no sea posible obtener una muestra de la matriz que no contenga analito, por ejemplo, es improbable que se consiga una muestra de residuos fotográficos sin plata. De esto se desprende que todas las mediciones analíticas, incluyendo la gráfica de calibración, de alguna manera deben realizarse con la propia muestra. Esto se logra en la práctica usando el **método de adiciones estándar**, el cual se utiliza mucho en la espectrometría de absorción atómica y emisión y también se aplica en el análisis electroquímico y en otras áreas. Se toman volúmenes iguales de solución problema, pero todas salvo una son “tratadas” por separado con cantidades conocidas y diferentes del analito, y *todas* se diluyen al mismo volumen. Después se determinan las señales instrumentales para todas estas soluciones y los resultados se representan como se muestra en la figura 5.9. Como de costumbre, la señal se representa en el eje y; mientras que en el eje x la escala se expresa en términos de las cantidades de analito *añadidas* (ya sea como peso absoluto o como concentración). La recta de regresión (no ponderada) se calcula de la manera usual, si bien el espacio ocupado por ella se extrapolará al punto sobre el eje x donde  $y = 0$ . Está claro que esta intersección negativa sobre el eje x corresponde a la cantidad de analito en la muestra del problema. El análisis de la figura muestra que este valor está dado por  $a/b$ , la razón de la ordenada en el origen y la pendiente de la recta de regresión. Ya que  $a$  y  $b$  están sujetas a error (véase la Sec. 5.5), el valor calculado también estará obviamente sujeto a error. Sin embargo, en este caso la cantidad no se predice a partir de un único valor medido de  $y$ , de manera que la fórmula de la desviación estándar  $s_{x_E}$ , del valor extrapolado de  $x$  ( $x_E$ ) no es la misma que en la ecuación (5.9). En su lugar, se utiliza:

$$s_{x_E} = \frac{s_{y/x}}{b} \left\{ \frac{1}{n} + \frac{\bar{y}^2}{b^2 \sum_i (x_i - \bar{x})^2} \right\}^{\frac{1}{2}} \quad (5.12)$$

Al aumentar nuevamente el valor de  $n$ , se mejora la precisión de la cantidad estimada: en general, deberían usarse al menos seis puntos en un experimento de adiciones estándar. Además, la precisión se mejora maximizando  $\sum_i (x_i - \bar{x})^2$ , así las soluciones de calibración cubrirían, si fuera posible, un intervalo considerable. Los límites de confianza para  $x_E$  se determinan, como siempre, como  $x_E \pm t s_{x_E}$ .

**Ejemplo.** La concentración de plata en una muestra de desechos fotográficos se determinó por espectrometría de absorción atómica con el método de las adiciones estándar. Se obtuvieron los siguientes resultados.

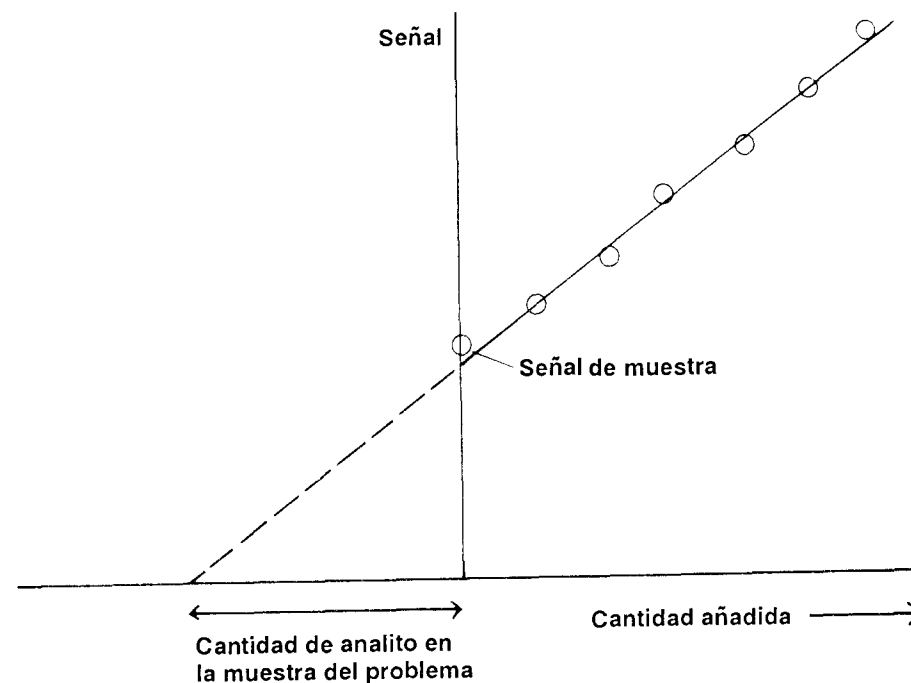


Figura 5.9 El método de las adiciones estándar.

Plata adicionada, $\mu\text{g}$ por ml de solución de muestra original	0	5	10	15	20	25	30
Absorbancia	0.32	0.41	0.52	0.60	0.70	0.77	0.89

Determine la concentración de plata en la muestra, y obtenga los límites de confianza al 95% para esta concentración.

Las ecuaciones (5.4) y (5.5) conduce a  $a = 0.3218$  y  $b = 0.0186$ . La razón entre estas cifras proporciona la concentración de plata en la muestra problema de  $17.3 \mu\text{g/ml}$ . Los límites de confianza para este resultado se pueden determinar con la ayuda de la ecuación (5.12). Aquí  $s_{y/x}$  es  $0.01094$ ,  $\bar{y} = 0.6014$ , y  $\sum (x_i - \bar{x})^2 = 700$ . El valor de  $s_{x_E}$  es entonces  $0.749$  y los límites de confianza son  $17.3 \pm 2.57 \times 0.749$ , es decir  $17.3 \pm 1.9 \mu\text{g/ml}$ .

Aunque es una aproximación elegante al problema común de efectos de interferencia por la matriz, el método de las adiciones estándar tienen varias desventajas: es difícil de automatizar, y puede utilizar cantidades de muestra más grandes que en otros métodos. En términos estadísticos, su principal desventaja es que es un método de extrapolación, por lo tanto es menos preciso que las técnicas de interpolación. En el ejemplo anterior podemos ver que, si una concentración desconocida de plata adicionada a la muestra de desechos fotográficos proporcionase un valor de absorbancia de  $0.65$ , la concentración de la plata adicionada debería ser  $17.6 \mu\text{g/ml}$ , con límites de confianza de  $17.6 \pm 1.6 \mu\text{g/ml}$ . Si comparamos este resultado con el obtenido en el ejemplo, es evidente la módica mejora en los límites de confianza. Como hemos visto, los límites de confianza para un punto sobre

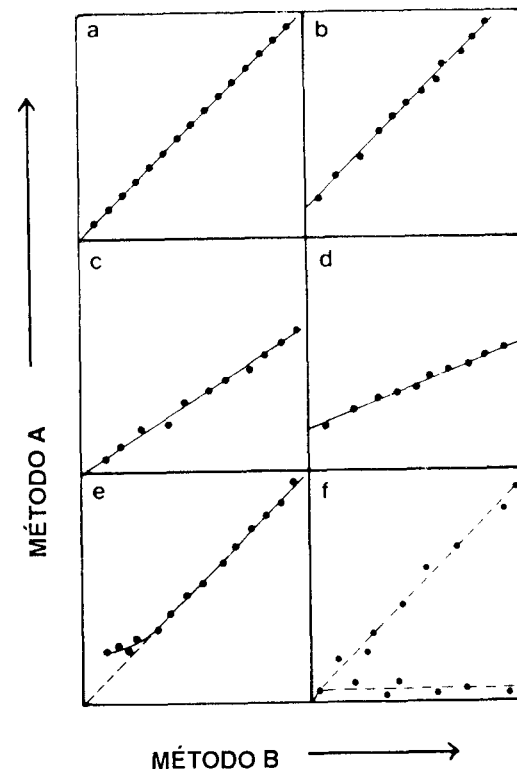
una recta de regresión varían con el valor de  $y$ , y alcanzan un mínimo cuando  $y = \bar{y}$  (véase la Sec. 5.6). En la práctica, como se muestra en este ejemplo y en el de la sección 5.6, esta variación en los límites de confianza con el cambio en  $y$  probablemente no es muy grande, de manera que los límites de confianza no son significativamente malos incluso para un resultado derivado de una extrapolación sustancial.

### 5.9 EL USO DE RECTAS DE REGRESIÓN PARA COMPARAR MÉTODOS ANALÍTICOS

Si un químico analítico desarrolla un método nuevo para determinar un analito concreto, deseará validarlo (entre otras técnicas) aplicándolo a un conjunto de muestras ya estudiadas mediante otro procedimiento estándar o de prestigio. Al realizar tal comparación, el principal interés será la identificación de errores sistemáticos (¿proporciona el nuevo método resultados significativamente más altos o más bajos que el procedimiento establecido?). En los casos donde un análisis se repite varias veces sobre un intervalo de concentración muy limitado, tal comparación puede hacerse usando las pruebas estadísticas descritas en un capítulo anterior (véanse las Seccs. 3.2 y 3.3). Dichos procedimientos no serán adecuados en análisis instrumental, los cuales a menudo se utilizan sobre intervalos de concentración grandes.

Cuando se comparan dos métodos a diferentes concentraciones de analito, se adopta normalmente el procedimiento que ilustra la figura 5.10. Uno de los ejes de la gráfica de regresión se utiliza para los resultados obtenidos por el nuevo método, y el otro para los que se obtienen aplicando el de comparación o referencia a las mismas muestras. (La importancia de asignar un eje a cada método se analiza posteriormente.) Así cada punto de la gráfica representa una única muestra analizada por dos métodos separados. Los métodos de las secciones precedentes se han aplicado para calcular la pendiente ( $b$ ), la ordenada en el origen ( $a$ ) y el coeficiente de correlación ( $r$ ) de la recta de regresión. Queda claro que si cada muestra conduce a un resultado idéntico con ambos métodos analíticos, la recta de regresión tendrá una ordenada en el origen de cero, y una pendiente y un coeficiente de correlación de 1 (véase la Fig. 5.10a). En la práctica, por supuesto, esto nunca ocurre incluso cuando los errores sistemáticos están completamente ausentes, los aleatorios aseguran que los dos procedimientos analíticos no darán resultados que concuerden exactamente en todas la muestras.

Pueden ocurrir desviaciones de la situación “ideal” ( $a = 0$ ,  $b = r = 1$ ) en una serie de circunstancias diferentes. En primer lugar, es posible que la recta de regresión tenga una pendiente de 1, pero no una ordenada en el origen de cero; o sea que un método de análisis puede producir un resultado más alto o más bajo que el otro en una cantidad fija. Tal error podría ocurrir si se calculó erróneamente la señal de fondo de uno de los métodos (véase la Fig. 5.10b). Una segunda posibilidad es que la pendiente de la recta de regresión sea  $> 1$  o  $< 1$ , lo que indica que puede darse un error sistemático en la pendiente de una de las gráficas de calibración individuales (Fig. 5.10c). Desde luego, estos dos errores pueden ocurrir simultáneamente (Fig. 5.10d). Otros tipos posibles de error sistemático se manifiestan si la representación gráfica es una curva (Fig. 5.10e). Problemas de especiación pueden proporcionar resultados sorprendentes (Fig. 5.10f). Este tipo de representación gráfica podría surgir si un analito se encuentra en dos formas químicas distintas, en proporciones que varían de una muestra a otra. Uno de los métodos (aquí representado en el eje



**Figura 5.10** El uso de una recta de regresión para comparar dos métodos analíticos: (a) muestra una concordancia perfecta entre los métodos para todas las muestras; de (b) a (f) ilustra los resultados de varios tipos de errores sistemáticos (véase texto).

y) podría detectar sólo una forma del analito, mientras que el segundo detectaría ambas formas.

En la práctica, el analista con frecuencia desea realizar una prueba para contrastar una ordenada en el origen significativamente diferente de cero, y una pendiente significativamente diferente de 1. Tales pruebas se realizan determinando los límites de confianza para  $a$  y  $b$ , generalmente al 95% de nivel de significación. De este modo, el cálculo es muy similar al descrito en la sección 5.5, y es más útil realizarlo utilizando un programa de computador sencillo.

**Ejemplo.** El nivel de plomo de diez muestras de jugo de fruta de determinó por un nuevo método de análisis potenciométrico de redisolución (*stripping*) (APR) empleando un electrodo de trabajo de carbón vitrificado, y los resultados fueron comparados con los obtenidos mediante la técnica de espectrometría de absorción atómica sin llama (EAA). Se obtuvieron los siguientes datos (todo los resultados en  $\mu\text{g/ml}$ ).

Muestra	1	2	3	4	5	6	7	8	9	10
Resultados EAA	35	75	75	80	125	205	205	215	240	350
Resultados APR	35	70	80	80	120	200	220	200	250	330

(Mannino, S., *Analyst*, 1982, **107**, 1466).

Estos resultados se representan en una línea de regresión (véase la Fig. 5.11), con los obtenidos por EAA en el eje  $x$  y los que resultaron por APR en el eje  $y$ . Usando los métodos de las secciones precedentes puede observarse fácilmente que:

$$a = 3.87, b = 0.968; r = 0.9945$$

Cálculos adicionales muestran que:

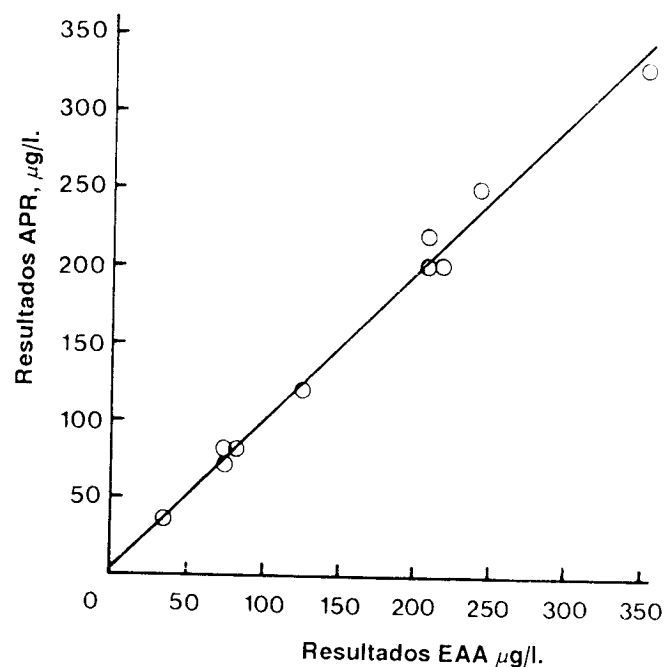
$$s_{y/x} = 10.56; s_a = 6.64; s_b = 0.0357$$

usando valores de  $t$  adecuados para 8 grados de libertad ( $t = 2.31$ ) dan los siguientes límites de confianza al 95% para la pendiente y ordenada en el origen:

$$a = 3.87 \pm 15.34 \text{ y } b = 0.963 \pm 0.083$$

De estos resultados se desprende que la pendiente y ordenada en el origen calculadas no difieren en forma significativa de los valores "ideales" de 1 y 0 respectivamente, y por lo tanto no existen diferencias sistemáticas entre las dos series de resultados.

En relación con este ejemplo pueden mencionarse otros dos aspectos. En primer lugar, la revisión de la bibliografía de química analítica muestra que los autores se preocupan con frecuencia de los valores del coeficiente de correlación en estudios comparativos; sin embargo, en este ejemplo no juega un papel directo al establecer si ocurren o no errores



**Figura 5.11** Comparación de dos métodos analíticos: la representación muestra los datos de la pág. 93.

sistemáticos. Incluso si las rectas de regresión son ligeramente curvadas, el coeficiente de correlación podría haber sido 1 (véase la Sec. 5.3). Es evidente, por lo tanto, que el cálculo de  $r$  tiene menor importancia en el presente contexto que el establecimiento de límites de confianza para la pendiente y ordenada en el origen. En algunos casos puede encontrarse que el valor de  $r$  no es muy próximo a 1, aun cuando la pendiente y ordenada en el origen no sean significativamente diferentes de 1 y 0 respectivamente. Tal resultado sugeriría una precisión muy pobre para uno o ambos de los métodos bajo estudio. Por supuesto la precisión de los dos métodos puede determinarse y compararse por los métodos de los capítulos 2 y 3. En la práctica es preferible que esto se haga *antes* de representar la recta de regresión que compara los dos métodos, la razón de esto se explica posteriormente. El segundo aspecto es que, si bien es claramente deseable comparar los métodos en un intervalo de concentraciones tan amplio como sea posible, en la práctica quizá no sea factible obtener muestras reales con concentraciones de analito igualmente espaciadas en el intervalo. En el ejemplo dado, los jugos de fruta examinados tienden a agruparse con contenidos en plomo claramente similares, de manera que las muestras que contengan plomo entre ca. 130 y 200  $\mu\text{g/l}$  y entre 260 y 320  $\mu\text{g/l}$  no se podrían estudiar. También veremos esto más adelante.

Este método, aunque a nivel internacional se ha adoptado prácticamente en estudios comparativos de métodos instrumentales, está abierto a serias objeciones teóricas por varias razones. En primer lugar, se ha recalcado a través de todo este capítulo, que la recta de regresión de  $y$  sobre  $x$  se calcula bajo el supuesto de que hay errores despreciables en los valores de  $x$  y se supone que todos los errores ocurren en la dirección  $y$ . Mientras que esto es generalmente válido en la generación de una gráfica de calibración para un único método analítico, no puede justificarse cuando la recta de regresión se usa con propósitos de comparación. En tal caso se puede tomar como cierto que los errores aleatorios existirán en ambos métodos analíticos, es decir, en ambas direcciones de  $x$  y  $y$ . Esto sugeriría que las ecuaciones utilizadas en el cálculo de la propia recta de regresión pudieran no ser válidas. Sin embargo, ensayos prácticos y simulaciones muestran que la presente aproximación conduce a resultados sorprendentemente fiables, siempre que se cumplan tres condiciones.

1. El método más preciso se representa en el eje  $x$ : Por eso se realizan investigaciones preliminares sobre las precisiones de los dos métodos (véase lo expuesto anteriormente).
2. Un número razonable de puntos (al menos 10, como en el ejemplo anterior) se representan en la comparación. Ya que los límites de confianza se basan en  $n - 2$  grados de libertad, es importante evitar valores pequeños de  $n$ .
3. Los puntos experimentales deberían cubrir el intervalo de concentración de interés de una manera aproximadamente uniforme: como hemos visto, puede ser difícil que se cumpla este requisito en estudios comparativos sobre muestras reales.

Existe una segunda objeción teórica al utilizar la recta de regresión de  $y$  sobre  $x$  como se calculó en las secciones 5.4 y 5.5, en la comparación de dos métodos analíticos. Esta recta de regresión no sólo supone que los errores en la dirección  $x$  son cero, sino también que los errores en los valores de  $y$  son *constantes*, es decir que no varían con la concentración, y por lo tanto todos los puntos tienen igual ponderación cuando se calculan la pendiente y la ordenada en el origen. Este supuesto probablemente carezca de validez en la práctica. En muchos análisis, la desviación estándar relativa (coeficiente de variación)

es aproximadamente constante en un intervalo de concentraciones: el error absoluto aumenta de este modo con la concentración del analito, en vez de tener el mismo valor en todas las concentraciones. Por consiguiente las rectas de regresión "no ponderadas" tienen también una validez muy cuestionable en otras situaciones, es decir, cuando se aplican a gráficas de calibración para un único procedimiento analítico. En principio, se deberían utilizar en su lugar rectas de regresión ponderadas. Este concepto se desarrolla en la siguiente sección. Mientras tanto, puede reafirmarse que, a pesar de las objeciones teóricas, las rectas de regresión no ponderadas proporcionan información útil en estudios comparativos si se cumplen los requerimientos enunciados anteriormente.

## 5.10 RECTAS DE REGRESIÓN PONDERADAS

Los comentarios hechos en las secciones anteriores sobre cálculos de regresión convencionales o no ponderados indican que los ponderados quizás deberían adoptarse más frecuentemente. Los cálculos que conllevan la utilización de los métodos de regresión ponderados son sólo un poco más complicados que los de las secciones anteriores, y pueden realizarse fácilmente en un microcomputador, pero requieren información adicional de los errores que surgen en los diferentes niveles de concentración, o al menos la formulación de supuestos adicionales acerca de tales errores. Esto explica quizás por que los cálculos de regresión no ponderada se usan con menos frecuencia. En esta sección se esbozan las aplicaciones de los métodos de regresión ponderada. Se supone que la recta de regresión ponderada se utiliza para la determinación de un único analito y no para comparar dos métodos separados.

Consideremos con más detalle la simple situación que surge cuando el error en el cálculo de una regresión es aproximadamente proporcional a la concentración del analito. Cuando los errores en los diferentes puntos de la gráfica de calibración se expresan por "barras de error" (véase la Fig. 5.12), éstas se alargan a medida que aumenta la concentración. En este caso es evidente que debe calcularse la recta de regresión para dar una ponderación adicional a aquellos puntos donde las barras de error son más cortas: es más importante que la recta calculada pase cerca de tales puntos que de los que representan concentraciones más altas con los errores más grandes. Este resultado se logra cuando se da a cada punto una ponderación inversamente proporcional a la varianza correspondiente,  $s_i^2$ ; este procedimiento lógico es de aplicación general. Así, si los puntos individuales se denotan con  $(x_1, y_1)$ ,  $(x_2, y_2)$ , etc., como es habitual, y las correspondientes desviaciones estándar son  $s_1, s_1$ , etc., entonces las ponderaciones individuales,  $w_1, w_1$ , etc., están dadas por:

$$w_i = s_i^{-2} / (\sum_i s_i^{-2} / n) \quad (5.13)$$

Como se verá las ponderaciones se han puesto en una escala de tal manera que su suma sea igual al número de puntos sobre la representación gráfica: esto simplifica los subsiguientes cálculos. La pendiente y la ordenada en el origen de la recta de regresión están dadas por:

$$b = \frac{\sum_i w_i x_i y_i - n \bar{x}_w \bar{y}_w}{\sum_i w_i x_i^2 - n \bar{x}_w^2} \quad (5.14)$$

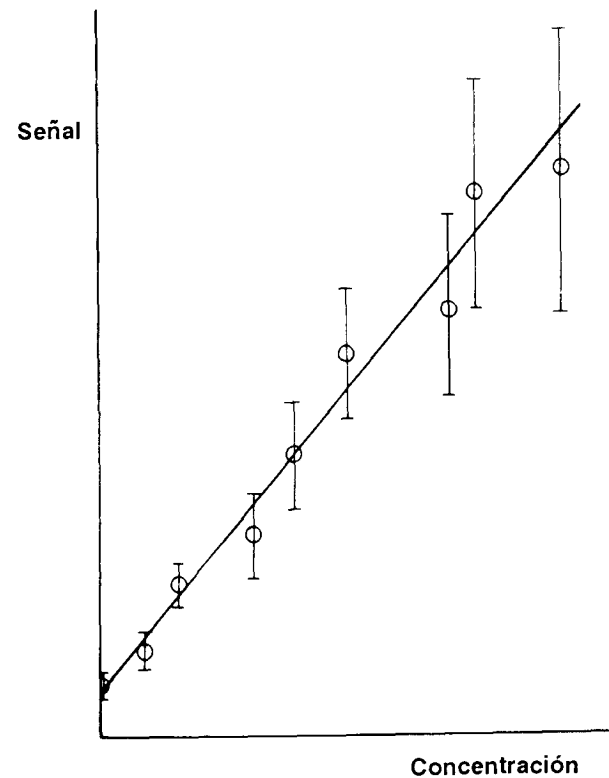


Figura 5.12 La ponderación de errores en un cálculo de regresión.

y

$$a = \bar{y}_w - b \bar{x}_w \quad (5.15)$$

En la ecuación (5.15),  $\bar{x}_w$  y  $\bar{y}_w$  representan las coordenadas del *centro de gravedad ponderado* ( $\bar{x}_w, \bar{y}_w$ ) por el que debe pasar la recta de regresión ponderada. Estas coordenadas están dadas por:  $\bar{x}_w = \sum_i w_i x_i / n$  y  $\bar{y}_w = \sum_i w_i y_i / n$ .

*Ejemplo.* Calcule las rectas de regresión ponderada y no ponderada para los siguientes datos de calibración. Calcule también para cada recta la concentración de las muestras problema con absorbancias de 0.100 y 0.600.

Concentración, $\mu\text{g/ml}$	0	2	4	6	8	10
Absorbancia	0.009	0.158	0.301	0.472	0.577	0.739
Desviación estándar	0.001	0.004	0.010	0.013	0.017	0.022

La aplicación de las ecuaciones (5.4) y (5.5) muestra que la pendiente y la ordenada en el origen de la recta de regresión *no ponderada* son respectivamente 0.0725 y 0.0133. Las concentraciones correspondientes a las absorbancias 0.100 y 0.600 son 1.20 y 8.09  $\mu\text{g/ml}$  respectivamente.



La recta de regresión *ponderada* es un poco más tediosa en cuanto a su cálculo: en ausencia de un programa de computador es habitual establecer una tabla como la que sigue.

$x_i$	$y_i$	$s_i$	$1/s_i^2$	$w_i$	$w_i x_i$	$w_i y_i$	$w_i x_i y_i$	$w_i x_i^2$
0	0.009	0.001	1000000	5.535	0	0.0498	0	0
2	0.158	0.004	62500	0.346	0.692	0.0547	0.1093	1.384
4	0.301	0.010	10000	0.055	0.220	0.0166	0.0662	0.880
6	0.472	0.013	5917	0.033	0.198	0.0156	0.0935	1.188
8	0.577	0.017	3460	0.019	0.152	0.0110	0.0877	1.216
10	0.739	0.022	2066	0.011	0.110	0.0081	0.0813	1.100
SUMAS:			1083943	5.999	1.372	0.1558	0.4380	5.768

De estos números queda claro que  $\bar{y}_w = 0.1558/6 = 0.0260$  y  $\bar{x}_w = 1.372/6 = 0.229$ . Mediante la ecuación (5.14),  $b$  se calcula a partir de

$$b = \frac{0.438 - (6 \times 0.229 \times 0.026)}{5.768 - [6 \times (0.229)^2]} = 0.0738$$

de manera que  $a$  está dado por  $0.0260 - (0.073 \times 0.229) = 0.0091$ .

Estos valores para  $a$  y  $b$  pueden utilizarse para mostrar que los valores de absorbancia de 0.100 y 0.600 corresponden a concentraciones de 1.23 y 8.01  $\mu\text{g/ml}$  respectivamente.

La cuidadosa comparación de los resultados de los cálculos de regresión ponderada y no ponderada es muy instructiva. Los efectos del proceso de ponderación son claros; el centro de gravedad ponderado ( $\bar{x}_w, \bar{y}_w$ ) está mucho más cerca del origen de la gráfica que el no ponderado ( $\bar{x}, \bar{y}$ ), y la ponderación dada a los puntos más próximos al origen y en particular al primer punto (0, 0.009) que tiene el menor de todos los errores asegura que la recta de regresión ponderada tiene una ordenada en el origen muy próxima a este punto. La pendiente y la ordenada en el origen de la recta ponderada son notablemente similares a las de la recta no ponderada, aunque los resultados de los dos métodos proporcionan valores muy parecidos para las concentraciones de muestras que tienen absorbancia de 0.100 y 0.600. No debe suponerse que esos valores semejantes surgen simplemente debido a que en este ejemplo los puntos experimentales se ajustan muy bien a una línea recta. En la práctica, las rectas de regresión ponderada y no ponderada derivadas de una serie de datos experimentales, tienen pendientes y ordenadas en el origen similares incluso si es relevante la dispersión de los puntos respecto a la recta.

Así, podríamos pensar que los cálculos de regresión ponderada son poco recomendables; requieren más información (en la forma de las estimaciones de la desviación estándar en varios puntos sobre la representación gráfica), y son más complejos de ejecutar, pero aparentemente proporcionan datos muy similares a los obtenidos a partir del método, mucho más simple, de regresión no ponderada. Dichas consideraciones de hecho pueden justificar que en la práctica se desprecien los cálculos de regresión ponderada. Pero un químico analítico que utiliza métodos instrumentales no emplea cálculos de regresión simplemente para determinar la pendiente y la ordenada en el origen de la gráfica de calibración y las concentraciones de muestras problema. También quiere obtener estimaciones de los errores o límites de confianza de esas concentraciones y en este contexto el

método de regresión ponderada proporciona resultados mucho más reales. En la sección 5.6 utilizamos la ecuación (5.9) para estimar la desviación estándar ( $s_{x_0}$ ) y por lo tanto los intervalos de confianza de una concentración calculada usando un solo valor de  $y$  y una recta de regresión no ponderada. Al aplicar esta ecuación a los datos del ejemplo anterior vemos que los límites de confianza no ponderados para las soluciones con absorbancias de 0.100 y 0.600 son  $1.20 \pm 0.65$  y  $8.09 \pm 0.63 \mu\text{g/ml}$  respectivamente. Como en el ejemplo de la sección 5.6, estos intervalos de confianza son muy similares. Sin embargo, en el ejemplo presente tal resultado no es completamente real; los datos experimentales muestran que los errores de los valores y observados, aumentan a medida que aumentan los propios valores de  $y$ ; es la situación esperada para un método que tenga una desviación estándar relativa aproximadamente constante. Podríamos esperar que este incremento en  $s_i$  con el de  $y$  también se reflejara en los límites de confianza de las concentraciones determinadas: los límites de confianza para la solución con una absorbancia de 0.600 serían mucho más grandes (es decir, peores) que los de la solución con una absorbancia de 0.100.

En los cálculos de regresión ponderados, la desviación estándar ( $s_{x_{0w}}$ ) de una concentración prevista está dada por:

$$s_{x_{0w}} = \frac{s_{(y/x)w}}{b} \left\{ \frac{1}{w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{b^2 (\sum_i w_i y_i^2 - n \bar{y}_w^2)} \right\} \quad (5.16)$$

En esta ecuación,  $s_{(y/x)w}$  está dada por:

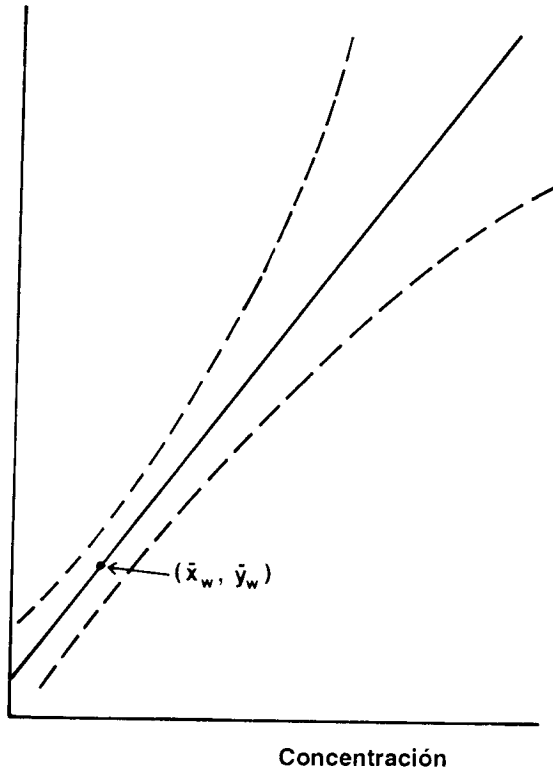
$$s_{(y/x)w} = \left\{ \frac{(\sum_i w_i y_i^2 - n \bar{y}_w^2) - b^2 (\sum_i w_i x_i^2 - n \bar{x}_w^2)}{n - 2} \right\}^{\frac{1}{2}} \quad (5.17)$$

y  $w_0$  es una ponderación adecuada para el valor de  $y_0$ . La ecuación (5.16) es claramente similar en su forma a la (5.9). Ello confirma que los puntos cercanos al origen, donde las ponderaciones son más altas, y los cercanos al centro de gravedad, donde  $y_0 - \bar{y}_w$  es pequeño, tendrán los límites de confianza más pequeños (véase la Fig. 5.13). La principal diferencia entre las dos ecuaciones es el término  $1/w_0$  en la ecuación (5.16). Ya que  $w_0$  disminuye bruscamente a medida que aumenta  $y$ , este término asegura que los límites de confianza aumentan cuando aumenta  $y_0$ , como esperábamos.

La aplicación de la ecuación (5.16) a los datos del ejemplo anterior muestra que las soluciones problema con absorbancias de 0.100 y 0.600 poseen intervalos de confianza para las concentraciones calculadas (1.23 y 8.01  $\mu\text{g/ml}$ ) de  $\pm 0.12$  y  $\pm 0.72 \mu\text{g/ml}$ , respectivamente. Vemos que estos intervalos de confianza son proporcionales a las absorbancias de las dos soluciones. Además, el intervalo de confianza para la muestra menos concentrada de las dos es más pequeño que en los cálculos de regresión no ponderados, mientras que para la muestra más concentrada ocurre lo contrario. Todos estos resultados concuerdan mucho más con el experimento de calibración real que con los resultados de los cálculos de regresión no ponderada.

Entonces podemos concluir que, si bien los cálculos de regresión ponderados son más complicados que sus homólogos no ponderados, dan resultados mucho más reales para los límites de confianza de concentraciones previstas en análisis instrumental convencional.

Señal



**Figura 5.13** Forma general de los límites de confianza para una concentración determinada utilizando una recta de regresión ponderada.

Además, los métodos de regresión ponderados pueden ser esenciales cuando se obtenga una recta por transformaciones algebraicas de una representación intrínsecamente curva (véase más adelante, Sec. 5.11). En la actualidad existen programas, para varios tipos de computadores personales, que realizan los cálculos de regresión ponderados, por lo que se debería fomentar más el uso de este método.

## 5.11 REGRESIÓN CURVILÍNEA

Hasta aquí, nuestro estudio de los métodos de calibración se ha centrado en experimentos de donde suponemos que la respuesta del instrumento es proporcional a la concentración de analito. Esto es usualmente válido, ya que los químicos analíticos utilizan tales métodos siempre que pueden, y toman precauciones experimentales para asegurar que la linealidad de la respuesta se conserve, en lo posible, en un amplio intervalo de concentraciones. Tales precauciones incluyen el control de la anchura de línea de emisión de una lámpara de cátodo hueco en espectrometría de absorción atómica, y la posición del porta muestras para minimizar los efectos del filtro interno en la espectrofluorimetría molecular. Sin embargo, a pesar de tales esfuerzos, muchos métodos analíticos proporcionan gráficas de calibración curvas en una parte del intervalo de concentración de interés. En concreto es muy usual la

situación en donde la gráfica de calibración es lineal (o aproximadamente lineal) a bajas concentraciones de analito, y sin embargo se vuelve curva a altas concentraciones. En otros casos, por ejemplo en inmunoensayos y en la respuesta de algunos electrodos selectivos de iones, la gráfica de calibración es obviamente curva en todas las concentraciones, y en este caso aún necesitamos respuestas a la lista de preguntas de la sección 5.2, si bien plantearán problemas estadísticos mucho más complicados que en el caso de experimentos de calibración lineales. La solución completa a estos problemas está fuera del alcance de este libro, aunque merece la pena darle cierta atención preliminar. El lector puede consultar la bibliografía para tratamientos más completos.

La primera pregunta es: ¿cómo detectamos la curvatura en una gráfica de calibración? Esto es, ¿cómo distinguimos entre una gráfica que se ajusta mejor por una línea recta, y una que se ajusta mejor por una curva poco pronunciada? Puesto que el grado de curvatura puede ser pequeño, y/o suceder sólo en parte de la gráfica, esto no es una pregunta trivial. Además, a pesar de su amplio uso para probar la bondad del ajuste de gráficas lineales, el coeficiente de correlación momento-producto ( $r$ ) es de poco valor al probar la curvatura: hemos visto (Sec. 5.3) que las líneas con una curvatura obvia todavía pueden dar valores de  $r$  muy elevados. Naturalmente, un analista esperaría que alguna prueba de curvatura se pudiera aplicar con bastante facilidad en trabajos de rutina sin cálculos extensos; algunas de estas pruebas están disponibles, y se basan en la utilización de los residuos de  $y$  y en las gráficas de calibración.

Como hemos visto (Sec. 5.5) un residuo de  $y$ ,  $y_i - \hat{y}_i$ , representa la diferencia entre un valor experimental de  $y$  y el valor calculado  $\hat{y}$  de la recta de regresión para el mismo valor de  $x$ . Si es apropiada una gráfica de calibración lineal, y si los errores aleatorios en los valores de  $y$  están distribuidos normalmente, los residuos en sí mismos deberían estar distribuidos normalmente en torno al valor cero. Si esto no ocurre en la práctica, entonces (despreciando la posibilidad de que los propios errores experimentales no estén distribuidos normalmente) debemos sospechar que la recta de regresión ajustada no es la correcta. En el desarrollo del ejemplo dado en la sección 5.5, los residuos  $y$  obtenidos fueron +0.58, -0.38, -0.24, -0.50, +0.34, +0.18 y +0.02; está claro que estos valores suman cero (teniendo en cuenta los posibles errores de redondeo, esto debe ser siempre verdadero), y aproximadamente están distribuidos en forma simétrica en torno a cero. Aunque es imposible tener la certeza —especialmente con un número pequeño de puntos— de que estos residuos estén normalmente distribuidos, tampoco existe evidencia en contra en este caso, es decir, no hay evidencia para justificar una gráfica de calibración no lineal.

Se sugiere una segunda prueba sobre la inspección de los signos de los residuos anteriormente expuestos. Si nos trasladamos a lo largo de la gráfica de calibración, es decir, cuando aumenta  $x$ , supondremos que existen residuos positivos y negativos en orden aleatorio si los datos están bien ajustados a una línea recta. Si, al contrario, intentamos ajustar una línea recta a una serie de puntos que caen realmente sobre una curva suave, entonces los signos de los residuos no tendrán un orden aleatorio más amplio, sino que se producirán *secuencias* de valores positivos y negativos. Examinando de nuevo los residuos expuestos anteriormente encontramos que el orden de los signos es + - - - - + +. Para probar si estas secuencias de residuos + y - indican la necesidad de una línea de regresión no lineal, necesitamos conocer la probabilidad de que tal ordenación pudiera ocurrir por azar; tales cálculos se describen en el próximo capítulo. Desafortunadamente debido al pequeño número de puntos es muy probable que éstas y otras secuencias pudieran realmente ocurrir por azar, de manera que cualquier conclusión que se extraiga debe tratarse

con gran precaución. Probablemente la elección entre métodos de regresión lineales y curvilínea sea más efectiva si se utilizan las técnicas de ajuste de curvas esbozadas posteriormente.

En la situación donde una gráfica de calibración sea lineal sobre parte de su intervalo, y curva en otra, es de gran importancia para el químico analítico establecer el intervalo sobre el que se supone linealidad. En el siguiente ejemplo se esbozan aproximaciones posibles a este problema.

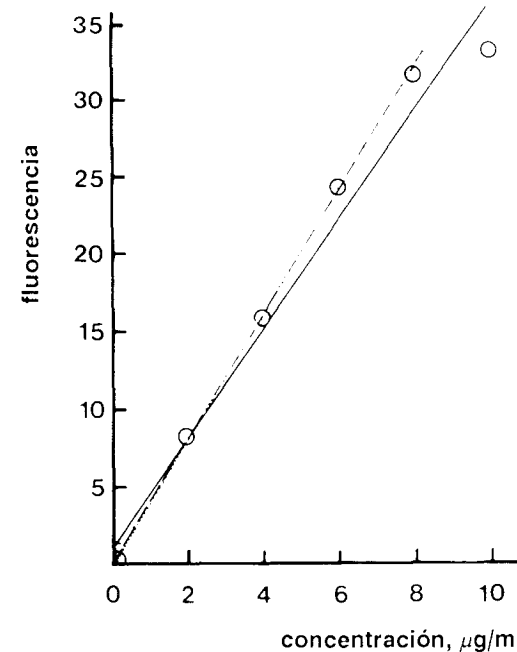
*Ejemplo.* Analice el intervalo de calibración lineal del siguiente experimento de fluorescencia.

Intensidad de fluorescencia	0.1	8.0	15.7	24.2	31.5	33.0
Concentración, $\mu\text{g/ml}$	0	2	4	6	8	10

El análisis de los datos muestra que la parte de la gráfica próxima al origen se adapta bastante bien a una línea recta con una ordenada en el origen próxima a cero y una pendiente cercana a 4. La fluorescencia de la solución estándar de 10  $\mu\text{g/ml}$  es claramente inferior a la esperada si se tiene en cuenta estas bases, y existe alguna posibilidad de que la desviación de la linealidad afecte también a la fluorescencia estándar de 8  $\mu\text{g/ml}$ . En primer lugar, se aplican cálculos de regresión lineal (no ponderada) a todos los datos; de la aplicación de los métodos de las secciones 5.3 y 5.4 se obtienen resultados  $a = 1.357$ ,  $b = 3.479$  y  $r = 0.9878$ . De nuevo tenga en cuenta que el valor alto de  $r$  puede ser engañoso, aunque se pueda utilizar en un sentido comparativo (véase más adelante). En este caso los residuos de  $y$  son  $-1.257$ ,  $-0.314$ ,  $+0.429$ ,  $+1.971$ ,  $+2.314$ , y  $-3.143$ , y la suma de los cuadrados de los residuos es igual a 20.981. La tendencia de los valores de los residuos sugiere que el último valor de la tabla está probablemente fuera del intervalo lineal.

Esta sospecha se confirma aplicando las ecuaciones de regresión lineal sólo a los cinco primeros puntos, resultando  $a = 0.100$ ,  $b = 3.950$  y  $r = 0.9998$ . La pendiente y la ordenada en el origen están mucho más cerca de los valores esperados para la parte de la gráfica más próxima al origen, y el valor de  $r$  es mayor que el primer cálculo. Los residuos de los primeros cinco puntos de esta segunda ecuación de regresión son 0, 0,  $-0.2$ ,  $+0.4$  y  $-0.2$ , con una suma de cuadrados de sólo 0.24. El uso de la segunda ecuación de regresión muestra que la fluorescencia esperada de un patrón de 10  $\mu\text{g/ml}$  es 39.6, es decir, el residuo es  $-6.6$ . El uso de la prueba  $t$  (Cap. 3) mostraría que este último residuo es significativamente más grande que el promedio de los otros: en forma alternativa se podría aplicar una prueba (Sec. 3.6) para demostrar que es un “residuo anómalo” entre los residuos. En este ejemplo, los cálculos, aunque tediosos, son necesarios: el gran residuo del último punto, unido con los residuos muy bajos de los otros cinco puntos y una suma de cuadrados enormemente reducida, confirma que el intervalo lineal del método no llega más allá de 10  $\mu\text{g/ml}$ .

Después de establecer que el último punto puede excluirse del intervalo lineal, podemos repetir el proceso para estudiar el punto (8, 31.5). Podemos realizarlo calculando la recta de regresión sólo para los primeros cuatro puntos de la tabla, con los resultados  $a = 0$ ,  $b = 4.00$ ,  $r = 0.9998$ . El valor del coeficiente de correlación sugiere que esta línea se ajusta tan bien a los puntos como la anterior, en la cual se usaban cinco puntos. Los residuos de este tercer cálculo son  $+0.1$ , 0,  $-0.3$ , y  $+0.2$ , con una suma de



**Figura 5.14** Regresión curvilínea: identificación del intervalo lineal. Los datos usados son los de la pág. 116; se muestran las rectas de regresión no ponderadas a través de todos los puntos (—), a través de los cinco primeros puntos (---), y a través de los cuatro primeros puntos (...).

cuadrados de 0.14. Con esta línea de calibración el residuo de  $y$  para la solución de 8  $\mu\text{g/ml}$  es  $-0.5$ : este valor es más grande que los otros residuos pero no en una cantidad significativa. De este modo podemos concluir que es razonablemente seguro incluir el punto (8, 31.5) dentro del intervalo lineal del método. Al tomar una decisión marginal de este tipo, el químico analítico tendrá en cuenta la exactitud requerida en sus resultados y el valor reducido de un método para el cual el intervalo de calibración es muy corto. Los cálculos descritos anteriormente se resumen en la figura 5.14.

Una vez que se ha tomado la decisión de que una serie de puntos de calibración no se ajustan satisfactoriamente a una línea recta, el analista puede jugar otra carta antes de resignarse a los complejos cálculos de regresión curvilínea. Puede ser capaz de **transformar** los datos de manera que una relación no lineal cambie a una lineal. Un método particularmente común es representar  $\log y$  y/o  $\log x$  en vez de  $y$  o  $x$ . Esto genera relaciones lineales de curvas originales de la forma  $y = px^q$ . Tales transformaciones se aplican muy regularmente a los resultados de ciertos métodos, por ejemplo inmunoensayos. Es importante señalar que la transformación puede afectar también a la naturaleza de los errores en los diferentes puntos de la gráfica de calibración. Suponga que en una serie de mediciones de la forma  $y = px^q$ , las magnitudes de los errores aleatorios en  $y$  son independientes de  $x$ ; la transformación de los datos en forma lineal al tomar logaritmos generará datos cuyos errores en  $\log y$  y no son independientes de los  $x$ . En este caso, y en cualquier otro donde la forma esperada de la ecuación se conozca a partir de consideraciones teóricas o de experiencia de mucho tiempo, es posible aplicar ecuaciones de regresión *ponderadas*

(véase la Sec. 5.10) a los datos transformados. Se puede demostrar que, si los datos de la forma general  $y = f(x)$  se transforman en una ecuación lineal  $Y = BX + A$ , el factor de ponderación,  $w_i$ , utilizado en las ecuaciones (5.14) a (5.17) se obtiene de la relación:

$$w_i = (1 / [dY_i / dy_i])^2 \quad (5.18)$$

En este ejemplo,  $y = px^q$ , de manera que  $Y = \ln y$ , y  $dY_i / dy_i = d \ln y_i / dy_i = 1/y_i$ . Así,  $w_i = y_i^2$ . Desafortunadamente, no existen muchos casos de química analítica donde la forma matemática exacta de una ecuación de regresión no lineal se conozca con certeza, de modo que esta aproximación tiene un valor restringido.

También debería considerarse que, en contraste con la situación descrita en el párrafo anterior, los resultados se pueden transformar para generar datos que podrían tratarse por métodos *no ponderados*. Datos de la forma  $y = bx$ , con errores en la dirección y claramente dependientes de  $x$  están sujetos algunas veces a transformaciones log-log: los errores en  $\log y$  tienen entonces una variación sensiblemente menor con  $\log x$ , de manera que los datos transformados se pueden estudiar mediante ecuaciones de regresión no ponderadas.

Si una simple transformación algebraica no puede convertir los datos experimentales en forma lineal, será necesario ajustar una curva al conjunto de puntos de calibración, y después utilizar esta curva para determinar la concentración de muestras problema. El ajuste de una curva puede ser un proceso iterativo complejo, en el cual los programas de computadores interactivos que muestren los resultados de las iteraciones sucesivas serán de gran ayuda para el analista. Una estrategia frecuente es tratar de ajustar una curva de expresión general  $y = a + bx + cx^2 + dx^3 + \dots$ , es decir, un polinomio en  $x$ . El éxito del proceso de ajuste a una curva puede valorarse, precisamente como en los cálculos de regresión lineal, mediante el método de los mínimos cuadrados, el programa del computador buscará la curva que minimice la suma de los cuadrados de los residuos de  $y$ . Se dispone de métodos no lineales ponderados, con un inevitable incremento en la complejidad. Todos los métodos de ajuste de curvas deben aplicarse con sentido común, si se tiene  $n$  puntos en la gráfica de calibración, el polinomio de orden más alto que puede utilizarse es el de orden  $(n - 1)$ . En general, se adoptará el polinomio de orden más bajo que proporciona un buen ajuste, y frecuentemente será satisfactoria una ecuación cuadrática o cúbica. Es muy importante señalar, sin embargo, que las curvas de calibración encontradas en la práctica se obtienen a menudo de la superposición de dos o más fenómenos físicos. En fluorimetría, por ejemplo, las representaciones señal contra concentración pueden ser aproximadamente lineales en solución muy diluida, pero a altas concentraciones exhibirán curvatura debido a (a) el tipo de óptica utilizada y (b) la falla de supuestos algebraicos sobre los que se pronostique una representación lineal. Los efectos (a) y (b) son independientes, de manera que pueden aparecer en la práctica muchas curvas de diferente forma. (Por razones de este tipo las curvas de forma conocida y pronosticable rara vez se encuentran en el trabajo analítico, véase lo anterior). Esto es tan importante que las representaciones de regresión no lineales no son *jamás* extrapolables más allá de las regiones exploradas realmente en el proceso de ajuste de la curva.

Si las gráficas de calibración no lineal son de hecho el resultado de la superposición de dos o más procesos fisico-químicos, parece oportuno sugerir que no debemos esperar ecuaciones algebraicas sencillas que ajusten los puntos de calibración con plena satisfacción. ¿Por qué no intentar ajustar los puntos de una curva que conste de varias secciones unidas con diferente forma matemática? Esta es exactamente la aproximación utilizada en

la actualidad con notable éxito a través de la aplicación de **funciones a trozos (splines)**. En la práctica, se usan más las **splines cúbicas**, de manera que la aproximación consiste en ajustar una curva mediante una serie de ecuaciones cúbicas. Estas "partes" cúbicas están unidas en  $p$  puntos llamados "nudos", y es esencial que las dos primeras derivadas de los dos trozos coincidentes en cada nudo sean iguales. Así pues, una **spline** cúbica completa tiene  $(p - 1)$  trozos o segmentos, cada uno con cuatro coeficientes (es decir, de la forma  $y = a + bx + cx^2 + dx^3$ ). Es necesario calcular  $(p - 1)(m + 1)$  coeficientes donde  $m$  sea el orden del polinomio, tres en este caso. Las estrategias utilizadas en estos cálculos son demasiado complejas para que las consideremos aquí, pero es notable que se hayan aplicado con éxito funciones de **spline** a ensayos de competencia obligatoria como radioinmunoensayos, a espectrometría de emisión atómica, y a experimentos de cromatografía gas-líquido.

Si la curvatura de la gráfica de regresión no es demasiado severa, y si los puntos de calibración no están demasiado separados (condiciones usualmente satisfechas en trabajos analíticos prácticos), puede usarse un método sencillo, aunque aproximado, en lugar de la aproximación compleja del ajuste de curvas. Esto conlleva el trazo de una línea recta entre cada par de puntos y el cálculo de concentraciones utilizando la interpolación lineal, es decir, la curva se trata como una serie de segmentos cortos y rectos. Este método no proporciona, por supuesto, límites de confianza para las concentraciones calculadas, pero su sencillez lo hace muy atractivo y los errores sistemáticos son, a menudo, sorprendentemente pequeños. Este punto se aclara en uno de los ejercicios al final del capítulo.

## BIBLIOGRAFÍA

- Acton, F. S., *Analysis of Straight Line Data*, Nueva York, Dover, 1966. Un tratamiento general bien escrito y extenso, con abundancia de ejemplos, incluyendo algunos de química. A pesar de su título, también hace referencia a la regresión curvilínea.
- Caulcutt, R. y Boddy, R., *Statistics for Analytical Chemists*, Londres, Chapman & Hall, 1983. Una aproximación totalmente pragmática, falto de teoría estadística, pero lleno de ejemplos de sentido común. Orientado hacia la química analítica industrial.
- Davies, O. L. y Goldsmith, P. L., *Statistical Methods in Research and Production*, Londres, Longmans, 1982. Un tratamiento bastante completo de los métodos de regresión y correlación, que pone un considerable énfasis sobre conceptos avanzados tales como regresión curvilínea y problemas de regresión múltiple.
- Draper, N. R. y Smith, H., *Applied Regression Analysis*, Nueva York, Wiley, 1966. Un trabajo clásico en este campo, con una cobertura de máximo alcance.
- Kaiser, H. y Menzies, A. C., *The Limit of Detection of a Complete Analytical Procedure*, Londres, Hilger, 1968. Una interesante discusión sobre límites de detección.
- Snedecor, G. W. y Cochran, W. G., *Statistical Methods*, Iowa State University, 1967. Aporta una excelente cantidad de procedimientos de regresión y correlación.
- Liteanu, C. y Ríca, I., *Statistical Theory and Methodology of Trace Analysis*, Chichester, Horwood, 1980. Amplia cobertura de métodos de análisis de trazas.

## EJERCICIOS

1. En un laboratorio provisto de equipo polarográfico se tomaron seis muestras de polvo a varias distancias del polarógrafo y se determinó el contenido de mercurio de cada muestra. Se obtuvieron los siguientes resultados.

Distancia al polarógrafo, m:	1.4	3.8	7.5	10.2	11.7	15.0
Concentración de mercurio, ng/g:	2.4	2.5	1.3	1.3	0.7	1.2

Examine la posibilidad de que la contaminación de mercurio surja del polarógrafo.

2. La respuesta de un ensayo colorimétrico para glucosa (GLU) se controló con la ayuda de soluciones estándar de glucosa. Determine el coeficiente de correlación a partir de los siguientes datos, y comente los resultados.

Concentración de GLU, mM:	0	2	4	6	8	10
Absorbancia	0.002	0.150	0.294	0.434	0.570	0.704

3. Se obtuvieron los siguientes resultados cuando se analizaron una serie de soluciones estándar de plata por espectrometría de absorción atómica de llama.

Concentración ng/ml:	0	5	10	15	20	25	30
Absorbancia:	0.003	0.127	0.251	0.390	0.498	0.625	0.763

Determine la pendiente y la ordenada en el origen de la gráfica de calibración, junto con sus límites de confianza.

4. Utilizando los datos del ejercicio 3, estime los límites de confianza para las concentraciones de plata en: (a) una muestra que proporciona una absorbancia de 0.308, 0.314, 0.347 y 0.312 en cuatro análisis separados.

5. Estime el límite de detección del análisis de plata de los datos del ejercicio 3.

6. El contenido de oro de una muestra de agua de mar concentrada se determinó por espectrometría de absorción atómica mediante el método de las adiciones estándar. Los resultados obtenidos fueron los siguientes.

Oro añadido, ng por ml de muestra concentrada:	0	10	20	30	40	50	60	70
Absorbancia:	0.257	0.314	0.364	0.413	0.468	0.528	0.574	0.635

Estime la concentración del oro en la muestra de agua de mar, y determine los límites de confianza para esta concentración.

7. La fluorescencia de una serie de soluciones ácidas de quinina fue determinada 5 veces; los resultados se exponen a continuación:

Concentración, ng/ml:	0	10	20	30	40	50
Intensidad de fluorescencia:	4	22	44	60	75	104
(unidades arbitrarias)	3	20	46	63	81	109
	4	21	45	60	79	107
	5	22	44	63	78	101
	4	21	44	63	77	105

Determine la pendiente y ordenada en el origen (a) la recta de regresión no ponderada y (b) la recta de regresión ponderada. Calcule, utilizando ambas rectas de regresión, los límites de confianza para la concentración de las disoluciones con intensidades de fluorescencia de 15 y 90 unidades.

8. Una determinación con un electrodo selectivo de iones (ESI) del sulfato procedente de sulfato reducido por bacterias se comparó con una determinación gravimétrica. Los resultados obtenidos se expresaron en miligramos de sulfuro.

Muestra:	1	2	3	4	5	6	7	8	9	10
Método ESI:	108	12	152	3	106	11	128	12	160	128
Gravimetría:	105	16	113	0	108	11	141	11	182	118

Comente la conveniencia del método ESI para esta determinación de sulfuro. (Al-Hitti, I. K., Moody, G. J. y Thomas, J. D. R., *Analyst*, 1983, **108**, 43.)

9. En la determinación de plomo en solución acuosa por espectrometría de absorción atómica con atomización en cámara de grafito, se obtuvieron los siguientes resultados.

Concentración de plomo, ng/ml:	10	25	50	100	200	300
Absorbancia:	0.05	0.17	0.32	0.60	1.07	1.40

Examine el intervalo de calibración lineal de este experimento. (Basado en Giri, S.K., Shields, C.K., Littlejohn, D. y Ottaway, J.M., *Analyst*, 1983, **108**, 244.)

10. En un análisis instrumental se obtuvieron los siguientes datos (unidades arbitrarias).

Concentración:	0	1	2	3	4	5	6	7	8	9	10
Señal:	0.2	3.6	7.5	11.5	15.0	17.0	20.4	22.7	25.9	27.6	30.2

Trace la gráfica de calibración, y compare las concentraciones correspondientes a las señales de 5, 16 y 27 unidades, (a) ajustando una línea recta a los puntos, (b) ajustando la curva  $y = 4x - 0.1x^2$ , (c) tratando la curva como una serie de segmentos lineales entre  $(x_1, y_1)$  y  $(x_2, y_2)$ ,  $(x_2, y_2)$  y  $(x_3, y_3)$ , etc., y (d) tratando la curva como una serie de segmentos lineales utilizando sólo puntos  $(x_1, y_1)$ ,  $(x_3, y_3)$ ,  $(x_5, y_5)$ , etcétera.

## Métodos no paramétricos y de ejecución rápida

### 6.1 INTRODUCCIÓN

Este capítulo describe una serie de métodos estadísticos útiles que difieren en un aspecto importante de los discutidos hasta ahora. Todas las pruebas desarrolladas en los capítulos anteriores suponen que los datos que se analizan siguen la distribución normal (Gaussiana). La validez de esta hipótesis se basa en el teorema del límite central (véase el Cap. 2), que demuestra que la distribución muestral de la media puede ser aproximadamente normal aunque la población de referencia tenga una distribución muy diferente: la aproximación mejora a medida que el tamaño de la muestra aumenta.

Sin embargo, existen varias razones por las cuales deberíamos estar interesados en métodos que no requieran tal hipótesis. En primer lugar, algunos conjuntos de datos que interesan a los químicos analíticos no se distribuyen normalmente. Por ejemplo (véase el Cap. 2), las concentraciones de anticuerpos en el suero sanguíneo de un grupo de personas diferentes puede expresarse aproximadamente como una distribución log-normal: dichos resultados no son en absoluto infrecuentes cuando se realiza una medición concreta sobre cada elemento de un grupo humano o de animales. En segundo lugar, en muchos experimentos utilizamos de hecho muestras pequeñas de datos (p. ej. sólo 3 ó 4 titulaciones en un análisis volumétrico), de manera que incluso el teorema del límite central sería de escasa relevancia. Finalmente, los métodos de los capítulos precedentes exigen, a menudo, muchos cálculos detallados; en contraposición, muchos de los métodos que se describen en este capítulo simplifican en gran parte, los cálculos. De hecho, algunas de las pruebas se realizan mentalmente con suma facilidad, característica muy deseable para efectuar una prueba.

Los métodos estadísticos que no suponen una forma específica de la distribución de la cual se toman los datos, se denominan métodos **no paramétricos** o de distribución libre. En las siguientes secciones se dará una idea general de una serie de métodos no paramétricos, y en muchos de ellos sólo se describirán las operaciones de las pruebas sin entrar en un desarrollo teórico detallado.

### 6.2 LA MEDIANA

En los capítulos anteriores se ha utilizado la media aritmética o promedio como la “medida de tendencia central” de un conjunto de resultados. Esto es bastante lógico cuando se supone la distribución (simétrica) normal, aunque en estadística no paramétrica se utiliza habitualmente en su lugar la **mediana**. Para calcular la mediana de  $n$  observaciones, se disponen en orden creciente (en el caso improbable de que  $n$  sea muy grande, este proceso de ordenación puede realizarse rápidamente con programas sencillos que se encuentran disponibles para la mayoría de los microcomputadores). Entonces, la mediana se define como el valor de la observación que ocupa el lugar  $\frac{1}{2}(n + 1)$  si  $n$  es impar, y el promedio de las observaciones que ocupan los lugares  $\frac{1}{2}n$  y  $(\frac{1}{2}n + 1)$  si  $n$  es par. La determinación de la mediana de una serie de resultados experimentales generalmente no requiere muchos cálculos. Además, en muchos casos puede ser una medida de tendencia central más realista que la media aritmética.

*Ejemplo.* Determine la media y la mediana para los cuatro valores siguientes de una titulación.

25.01    25.04    25.06    25.21 ml

Es fácil calcular que la media de estas cuatro observaciones es de 25.08 ml y que la mediana es de 25.05 ml —en este caso es el promedio del segundo y tercer valores, ya que las medidas están ordenadas numéricamente. La media es mayor que cualquiera de los tres valores más próximos entre sí (25.01, 25.04 y 25.06 ml) y de este modo puede ser una medida de tendencia central menos realista que la mediana. En lugar de calcular la mediana podríamos utilizar los métodos del capítulo 3 para probar el valor 25.21 como un posible resultado anómalo, y determinar la media de acuerdo con el resultado obtenido, pero esta aproximación implica cálculos extras y *supone* que los datos proceden de una distribución normal.

Este simple ejemplo ilustra una propiedad valiosa de la mediana: ésta no se ve afectada por los resultados extremos o anómalos. Los límites de confianza (véase el Cap. 2) para la mediana se pueden estimar con la ayuda de la distribución binomial. Este cálculo puede realizarse aun cuando el número de mediciones sea pequeño, pero no es habitualmente necesario en química analítica donde la mediana se utiliza por lo general sólo como una estimación rápida de un promedio. El lector puede consultar la bibliografía para mayor información.

En estadística no paramétrica la medida habitual de dispersión (que sustituye a la desviación estándar) es el **recorrido intercuartílico**. Como ya hemos visto, la mediana divide la muestra de mediciones en dos mitades iguales; si cada una de éstas se divide posteriormente en otras dos mitades, los puntos de división se denominan **cuartiles inferior** y **superior**. Al hacer este cálculo se utilizan varios convenios diferentes; de nuevo, el lector interesado puede consultar la bibliografía. El recorrido intercuartílico no es muy usado en el trabajo analítico, pero varias pruebas estadísticas se pueden ejecutar con él.

### 6.3 LA PRUEBA DE SIGNOS

La prueba de signos es uno de los métodos no paramétricos más simples y se empezó a estudiar por primera vez a principios del siglo XVIII. Se puede utilizar de diferentes formas, la más simple se ilustra con el siguiente ejemplo.

*Ejemplo.* Un preparado farmacéutico exige tener un contenido mediano del 8% de un constituyente concreto. En la práctica se encontraron lotes que contenían 7.3, 7.1, 7.9, 9.1, 8.0, 7.1, 6.8 y 7.3% del constituyente. Al nivel de significación del 5%, ¿indica esto que la mediana exigida es incorrecta?

En el capítulo 3 (véase la Sec. 3.2) se demostró que tales problemas podrían abordarse utilizando la prueba *t* después de calcular la media y la desviación estándar de los datos experimentales. La prueba *t* supone que los datos se distribuyen normalmente. La prueba de signos prescinde de tal hipótesis y es mucho más fácil de realizar. A cada valor experimental se le resta a su vez, la mediana postulada, y se considera el signo de cada resultado. Se ignoran *por completo* aquellos valores que son iguales a la mediana postulada. En este caso, por tanto, disponemos efectivamente de siete valores experimentales, seis de ellos inferiores a la mediana y a los que individualmente se les da el signo menos, y uno más grande que la mediana al que se le da el signo más. Para contrastar si esta preponderancia de signos menos es significativa utilizamos la ley binomial. Esta ley establece que la probabilidad de que aparezcan *r* signos menos entre *n* signos está dada por:

$$P(r) = {}^nC_r p^r q^{(n-r)} \quad (6.1)$$

donde  ${}^nC_r$  indica el número de combinaciones de *r* elementos de un total de *n* elementos, *P* es la probabilidad de que aparezca un signo menos en uno de los resultados y *q* es la probabilidad de que no aparezca un signo menos en uno de los resultados, es decir,  $q = 1 - p$ . Ya que la mediana se define de manera que la mitad de los resultados experimentales se sitúa por encima de ella y la otra mitad por debajo, queda claro que si la mediana es 8.0 en este caso, entonces *p* y *q* deberían ser  $\frac{1}{2}$ . Utilizando la ecuación (6.1) encontramos que  $P(6) = {}^7C_6 (\frac{1}{2})^6 \times \frac{1}{2} = 7/128$ . De manera similar podemos calcular que la probabilidad de obtener siete signos menos,  $P(7)$ , es  $1/128$ . Por lo tanto, en general la probabilidad de obtener *seis o más* signos negativos en nuestro experimento es  $8/128$ . La cuestión sólo consiste en saber si los datos difieren significativamente de la mediana postulada. Por consiguiente, debemos realizar una prueba de dos colas (véase el Cap. 3), es decir, debemos calcular la probabilidad de obtener 6 o más signos idénticos (o sea,  $\geq 6$  signos más o  $\geq 6$  signos menos), cuando se toman al azar siete resultados. Esto es claramente  $16/128 = 0.125$ . Luego comparamos esta cifra con 0.05, es decir, realizamos la prueba al nivel de significación del 5%. Puesto que el valor experimental es mayor que 0.05, nuestra hipótesis nula (es decir, que los datos proceden de una población de mediana 8.0) no se puede rechazar. Igual que en el capítulo 3, es importante observar que no se ha probado que los datos proceden de tal población; sólo se ha probado que tal hipótesis no se puede rechazar.

Queda claro a partir de este ejemplo que la prueba de signos conllevará al uso frecuente de la distribución binomial con  $p = q = \frac{1}{2}$ . Esta aproximación es tan común en la estadística no paramétrica que la mayoría de las tablas estadísticas incluyen los datos necesarios, lo cual permite que tales cálculos se realicen instantáneamente (véase la tabla A.7). Además,

en muchas situaciones prácticas, un analista utilizará siempre el mismo valor de *n*, es decir, tomará siempre el mismo número de lecturas o muestras, y podrá memorizar con facilidad las probabilidades correspondientes a los diversos signos + o -.

La prueba de signos puede utilizarse también como una alternativa no paramétrica a la prueba *t* por parejas (véase la Sec. 3.4) para comparar dos grupos de resultados para las mismas muestras. Así, si se examinan diez muestras con cada uno de los dos métodos, A y B, se puede contrastar si los dos proporcionan lecturas significativamente diferentes, calculando para cada muestra la diferencia en los resultados, es decir [(resultado obtenido por el método A) - (resultado obtenido por el método B)]. La hipótesis nula será que los dos métodos no proporcionan resultados significativamente diferentes —en la práctica esto significará nuevamente que la probabilidad de obtener un signo más (o un signo menos) es 0.5. Ahora se puede comparar el número de signos más o menos obtenidos con la probabilidad que se deriva de la ecuación (6.1). En los ejercicios que se encuentran al final de este capítulo aparece un ejemplo de esta aplicación de la prueba de signos.

Otro uso de la prueba de signos es indicar una tendencia. Esta aplicación se ilustra mediante el siguiente ejemplo.

*Ejemplo.* El nivel de una hormona en el plasma sanguíneo de un paciente se mide a la misma hora cada día durante diez días. Los datos resultantes son:

Día	1	2	3	4	5	6	7	8	9	10
Nivel, ng/ml	5.8	7.3	4.9	6.1	5.5	5.5	6.0	4.9	6.0	5.0

¿Existe alguna evidencia de que haya una tendencia en la concentración de hormona?

Utilizando métodos paramétricos, sería posible elaborar una gráfica de regresión lineal de tales datos y contrastar si su pendiente difiere significativamente de cero (véase el Cap. 5). Tal aproximación supondría que los errores se distribuyen normalmente y que cualquier tendencia que apareciese sería del tipo lineal. La aproximación no paramétrica es de nuevo más simple. Los datos se dividen en dos grupos iguales de la siguiente forma:

5.8	7.3	4.9	6.1	5.5
5.5	6.0	4.9	6.0	5.0

Al resultado del primer día se le resta el del sexto día, al del segundo día el del séptimo, etc. Los signos de las diferencias entre los pares de valores en las cinco columnas se determinan de esta manera y son +, +, 0, +, +. Como es habitual, el valor 0 se ignora, y quedan cuatro resultados, todos ellos positivos. La probabilidad de obtener cuatro signos idénticos en cuatro pruebas es claramente  $2 \times (1/16) = 0.125$ . (Observe que nuevamente se utiliza una prueba de dos colas, ya que la tendencia en el nivel de hormona podría ser ascendente o descendente.) La hipótesis nula, de que no existe ninguna tendencia en los resultados, no se puede rechazar en esta forma al nivel de significación del 5%.

El lector crítico puede encontrar este resultado bastante insatisfactorio. Si se observan los datos, es difícil resistirse a la conclusión de que, a pesar de los resultados de la prueba



de signos, hay de hecho una tendencia descendente en los resultados. Por ejemplo, la media de los tres primeros resultados es 6.00, la de los cuatro siguientes es 5.78, y la de los tres últimos es 5.30. Esta aparente discrepancia en parte es una consecuencia de la filosofía del contraste de hipótesis: no podemos rechazar la hipótesis de que no existe tendencia en los resultados, pero esto no significa que tal tendencia *deba necesariamente* estar ausente. Sin embargo, el ejemplo indica también que por la extrema sencillez de la prueba de los signos se pierde cierta potencia estadística. La prueba no utiliza toda la información ofrecida por los datos, por eso no es sorprendente encontrar que también proporciona menos información discriminante. En secciones posteriores se analizarán métodos no paramétricos que *utilizan* las magnitudes de los resultados individuales así como sus signos.

#### 6.4 LA PRUEBA DE RACHAS DE WALD-WOLFOWITZ

En algunos casos puede interesarnos no sólo si las observaciones generan signos positivos o negativos, sino también si éstos aparecen en una secuencia aleatoria. En la sección 5.11, por ejemplo, se demostró que si una línea recta es un buen ajuste a una serie de puntos de calibración, los residuos positivos o negativos ocurrirán más o menos al azar. Por el contrario, al intentar ajustar una línea recta a un conjunto de puntos que realmente se sitúan en una línea curva resultarán secuencias no aleatorias de signos positivos o negativos; podría ser, por ejemplo, una secuencia de signos +, seguida de una de signos -, y luego otra de signos +. Dichas secuencias se conocen técnicamente como **rachas** —el significado utilizado aquí es el mismo que cuando alguien se refiere a “una racha de mala suerte”, o cuando un deportista experimenta “una racha de malos resultados”. En el caso del ajuste de curvas, queda claro que una secuencia no aleatoria de signos + y - conducirá a un número más pequeño de rachas que una aleatoria. El método de Wald-Wolfowitz prueba si el número de rachas es suficientemente pequeño para que se rechace la hipótesis nula de una distribución aleatoria de los signos. El número de rachas en los datos experimentales se compara con los números de la tabla A.8, que se presentan al nivel de significación del 5%. En esta tabla se entra utilizando los valores apropiados de  $N$ , el número de signos +, y  $M$ , el número de signos -. Si el número experimental de rachas es *más pequeño* que el valor tabulado, entonces se puede rechazar la hipótesis nula.

*Ejemplo.* Para ajustar una recta a un conjunto de 12 puntos de calibración se proponen ecuaciones de regresión lineal. Los signos de los residuos resultantes ordenados por valores de  $x$  crecientes son: + + + + - - - - + +. Comente si sería mejor intentar ajustar una curva a los puntos.

Aquí es evidente que  $N = M = 6$ , y que el número de rachas es 3. La tabla A.8 muestra que, al nivel de significación del 5%, el número de rachas debe ser  $< 4$  si se rechaza la hipótesis nula. Así, en este caso, podemos rechazar la hipótesis nula y concluir que la secuencia de signos + y - no es aleatoria. Por lo tanto resulta poco satisfactorio intentar ajustar una línea recta a los puntos experimentales, y en su lugar se propone una gráfica de regresión curvilínea.

Existen otros dos puntos de interés en relación con la prueba de Wald-Wolfowitz. El primero es que merece la pena observar que se puede utilizar con cualesquiera resultados

que se puedan dividir o convertir en sólo dos categorías. Supongamos, por ejemplo, que se dispone de los tiempos de funcionamiento de doce lámparas que se han ido colocando sucesivamente en un espectrómetro como fuentes de luz y que son 450, 420, 500, 405, 390, 370, 380, 395, 370, 420, y 430 horas. El tiempo de vida mediano, es en este caso el promedio de los tiempos de funcionamiento de las lámparas sexta y séptima cuando los datos están dispuestos en orden creciente, es de 400 horas. Si a todas aquellas lámparas con tiempos de funcionamiento menores, que el mediano se les asigna un signo -, y a todas aquella con tiempos de vida superiores se les asigna un signo +, entonces resulta la siguiente frecuencia + + + + - - - - + +. Ésta es la misma secuencia que la del ejemplo de regresión anterior, y se demuestra de la misma manera que es significativamente no aleatoria. En este caso, las variaciones significativas que se producen en los tiempos de funcionamiento se podrían explicar porque las lámparas proceden de diferentes fabricantes o de diferentes lotes.

También es importante resaltar que puede incomodarnos mucho el tratar con números inusualmente grandes de rachas cortas, así como con números inusualmente pequeños de rachas largas. Así pues, si se dan 6 signos + y 6 signos - en el orden + - + - + - + - + -, podríamos sospechar que existe una secuencia no aleatoria. La tabla A.8 muestra que, con  $N = M = 6$ , un total de 11 ó 12 rachas indica que la hipótesis nula de un orden aleatorio se debería rechazar y sospechar una cierta periodicidad en los datos.

#### 6.5 PRUEBAS BASADAS EN EL RECORRIDO DE LOS RESULTADOS

En los capítulos anteriores la desviación estándar se ha utilizado como la medida de dispersión o “variabilidad” más común de un conjunto de resultados. En estadística no paramétrica se utiliza frecuentemente el recorrido intercuartílico (véase la Sec. 6.1) como una medida de dispersión. Además varias pruebas muy útiles se basan en el recorrido completo  $w$  de la muestra, es decir, la diferencia entre el valor más grande y el más pequeño. Es evidente que  $w$  será más fácil de calcular que el recorrido intercuartílico, y de hecho los cálculos de la mayoría de las pruebas descritas en esta sección se pueden realizar mentalmente. No obstante, debe recalarse que aunque estas pruebas son rápidas, *no son* no paramétricas: por ejemplo los cálculos utilizan la media aritmética de la muestra y no la mediana.

En lugar de las dos aplicaciones elementales de la prueba se pueden utilizar pruebas simples de recorridos. En la sección 3.2 se describió el uso de la prueba  $t$  para comparar una media experimental con un valor conocido o estándar. En lugar de esta prueba el estadístico  $T_i$  se puede calcular a partir de la ecuación

$$T_i = |\bar{x} - V|/w \quad (6.2)$$

donde  $\bar{x}$  es la media y  $V$  el valor estándar.

*Ejemplo.* En una método de determinación de mercurio por espectrometría de absorción atómica de vapor frío se obtuvieron los siguientes valores para un material estándar que contenía el 38.9% de mercurio:

38.9, 37.4 37.1%



¿Existe alguna evidencia de error sistemático? Observe que este ejemplo es el mismo que el de la sección 3.2.

Es evidente que la media,  $\bar{x}$ , en este caso es 37.8%, y el recorrido,  $w$ , es 1.8%. El valor de  $T_i$  es, por tanto, [ecuación (6.2)]  $|37.8 - 38.9| / 1.8 = 0.611$ . Este valor se compara con el de la tabla A.9, para  $n = 3$  y  $P = 0.05$ . El valor tabulado es 1.304: ya que el valor experimental es más pequeño no podemos rechazar la hipótesis nula, o sea, que los datos experimentales podrían venir de una población con un valor medio de 38.9% de mercurio. Este resultado —que al nivel de significación del 5% manifiesta ausencia de error sistemático— es el mismo que el obtenido en la sección 3.2 mediante la prueba  $t$ , aunque los cálculos en este caso se realizan obviamente con mayor facilidad. Hay que agregar que la prueba  $T_i$  puede utilizarse también para estimar límites de confianza; en este caso los límites de confianza estarán dados por  $\bar{x} \pm (1.304 \times 1.8) = 37.8 \pm 2.3\%$  (un recorrido que, como se esperaba, incluye el 38.9%).

Una prueba muy similar puede utilizarse en vez de la prueba  $t$  en la comparación de dos valores medios. Si se compara un conjunto de datos de media  $\bar{x}_1$  y recorrido  $w_1$  con un segundo conjunto de datos con el mismo valor de  $n$  y de media  $\bar{x}_2$  y recorrido  $w_2$  respectivamente, el estadístico  $T_d$  se calcula a partir de la expresión:

$$T_d = 2|\bar{x}_1 - \bar{x}_2| / (w_1 + w_2) \quad (6.3)$$

Como es habitual, si el valor experimental de  $T_d$  es menor que el valor tabulado (véase la tabla A.10), entonces se acepta la hipótesis nula de igualdad entre las dos medias muestrales. Un ejemplo de este tipo de cálculo se propone al final del capítulo. Otra prueba con el mismo propósito y que se ha hecho popular es la **prueba de recorridos de Lord**. En este caso el estadístico de la prueba,  $L$ , está dado por:

$$L = |\bar{x}_1 - \bar{x}_2| / (w_1 + w_2) \quad (6.4)$$

es decir utiliza la suma de recorridos en lugar del recorrido promedio. Estas pruebas son útiles cuando se desea comparar muestras pequeñas ( $n$  puede ser tan pequeño como 2), pero (como la prueba  $t$  descrita en la sección 3.3) suponen que la población de partida es normal, y que la dispersión es razonablemente similar en las dos muestras. De nuevo su ventaja radica en la sencillez de los cálculos. La tabla A.10 proporciona un conjunto de valores tabulados para  $L$ : el rechazo de la hipótesis nula se produce cuando surgen valores experimentales superiores a los valores tabulados.

Incluso se puede utilizar otras pruebas simples de recorrido en lugar de la prueba  $F$  cuando se compara la dispersión de dos conjuntos de resultados. En lugar de comparar varianzas, el nuevo estadístico,  $Fr$ , compara recorridos; esto es,  $Fr = w_1/w_2$  o  $w_2/w_1$ , el cociente que sea mayor que 1. El valor experimental  $Fr$  se compara, como es habitual, con los valores tabulados (véase la tabla A.11). Se rechaza la hipótesis nula de que las dos muestras proceden de poblaciones de varianzas iguales, si el valor experimental es mayor que el valor tabulado. Como en la prueba  $F$  convencional se dispone de versiones de una y dos colas de este “sustituto de la prueba  $F$ ”, la tabla A.11 se debe utilizar con cuidado. En los ejercicios al final del capítulo se propone un ejemplo del uso de este método.

Ya se utilizó el recorrido en otra prueba descrita en el capítulo 3 (véase la sección 3.6), la prueba  $Q$  para resultados anómalos. Entonces se demostró que el valor experimental de  $Q$  [dado por  $[(\text{valor sospechoso}) - (\text{valor más próximo})] / \text{recorrido}$ ] se puede comparar con un conjunto de valores tabulados (tabla A.4) para establecer si una medición sospechosa podría ser candidata al rechazo. Se puede demostrar que una forma más simple, pero aproximada de esta prueba, prescinde de tablas. Un resultado anómalo se puede rechazar al nivel de significación  $P = 0.05$  si  $Q > \sqrt{(2/n)}$ , y al nivel  $P = 0.01$  si  $Q > \sqrt{(3/n)}$ , donde  $n$  es el número de mediciones *incluyendo* aquella que es aparentemente anómala.

Esta prueba sirve para recordar que los métodos de recorridos presentan la gran desventaja de verse muy afectados por uno o más resultados anómalos. En contraposición, los métodos verdaderamente no paramétricos son por lo general inmunes a tales resultados. Las pruebas de recorrido poseen otra desventaja: tienen menos potencia estadística que los métodos correspondientes descritos en el capítulo 3. Debido a que las pruebas de recorrido no utilizan todos los datos disponibles, habrá casos en los que una prueba  $F$ , por ejemplo, indique que la hipótesis nula apropiada pueda ser rechazada, pero su alternativa más simple, el sustituto de la prueba  $F$ , indique que la hipótesis nula debería ser aceptada. La popularidad de las pruebas de recorrido indica que la simplicidad en los cálculos es una compensación más que suficiente a todos estos problemas.

## 6.6 LA PRUEBA DE RANGOS Y SIGNOS DE WILCOXON

La sección 6.3 describía el uso de la prueba de signos. Su importancia reside en los supuestos mínimos que se hacen sobre los datos experimentales. No se supone que la población de la cual se toma la muestra sea normal, ni incluso que sea simétrica —la única información *a priori* necesaria es el valor de la mediana. Una cierta desventaja de la prueba de signos es que no utiliza toda la información disponible. Sólo es necesario saber si una medición individual es más grande o más pequeña que la mediana— la magnitud de esta desviación no se utiliza en absoluto.

En muchos casos un analista tendrá suficientes razones para creer que sus mediciones se distribuyen *simétricamente* pero no desea suponer que siguen distribución normal. Este supuesto de datos simétricos, y la consecuencia de que la media y mediana de la población sean iguales, permite desarrollar pruebas de significación más potentes. Wilcoxon contribuyó con importantes avances a este respecto, y su prueba de rangos y signos tiene varias aplicaciones. Su mecanismo se ilustra mucho mejor con un ejemplo.

*Ejemplo.* Se encontró que los niveles de plomo en sangre (en pg/ml) de siete niños eran 104, 79, 98, 150, 87, 136 y 101. ¿Podrían proceder estos datos de una población que se supone simétrica, con mediana (media) de 95 pg/ml?

Los datos comparados con el valor de referencia (95) toman los valores

9, -16, 3, 55, -8, 41, 6

En primer lugar y prescindiendo del signo se ordenan estos valores de menor a mayor, resultando:

3, 6, 8, 9, 16, 41, 55

A continuación se incorporan sus signos (en la práctica estos dos pasos se harían de una vez):

3, 6, -8, 9, -16, 41, 55

Los números entonces **se jerarquizan**; en este proceso mantienen sus signos pero se les asignan números que indican su orden (o rango), así:

1, 2, -3, 4, -5, 6, 7

Los rangos positivos suman 20 y los negativos 8. La *menor* de estas dos cifras (8) se toma como el estadístico de la prueba. El teorema binominal dará la probabilidad de que aparezca este número. Si los datos proceden de una población con una mediana de 95, se esperaría que las sumas de los rangos positivos y negativos fuesen aproximadamente iguales; si la mediana de la población fuese muy diferente de 95, la suma de los rangos positivos y negativos sería diferente. La probabilidad de que aparezca una suma concreta en la práctica está dada por una serie de tablas (véase la tabla A.12). En esta prueba se rechaza la hipótesis nula si el valor tabulado es *menor o igual que* el valor experimental, es decir, la situación opuesta de la observada en la mayoría de las pruebas de significación. En este ejemplo, el examen de la tabla A.12 muestra que, para  $n = 7$ , el estadístico de la prueba debe ser menor o igual que 2 para que la hipótesis nula —que los datos *proceden* de una población con una mediana (media) de 95— se puede rechazar a un nivel de significación de  $P = 0.05$ . En este ejemplo, la hipótesis nula se debe aceptar claramente. Como de costumbre, se utiliza una prueba de dos colas, aunque quizás haya ocasiones en los que una prueba de una cola sea más adecuada.

Una ventaja importante de la prueba de rangos y signos es que también se puede utilizar para datos por parejas, ya que se pueden transformar en el tipo de datos dados en el ejemplo anterior. De esta forma se puede utilizar el método de rangos y signos como una alternativa no paramétrica a la prueba  $t$  por parejas (véase la Sec. 3.4).

*Ejemplo.* La siguiente tabla proporciona el porcentaje de concentración de zinc determinado por dos métodos diferentes, para cada una de las ocho muestras de alimentos.

Muestra	Titulación con AEDT	Espectrometría atómica
1	7.2	7.6
2	6.1	6.8
3	5.2	4.6
4	5.9	5.7
5	9.0	9.7
6	8.5	8.7
7	6.6	7.0
8	4.4	4.7

¿Hay alguna diferencia sistemática entre los resultados de los dos métodos?

La aproximación a este tipo de problemas es muy simple. Si no hay diferencia sistemática entre los dos métodos, entonces se esperaría que las diferencias entre los resultados para cada muestra, es decir, [(resultado volumétrico) - (resultado espectro-

métrico)]) se distribuirán simétricamente en torno a cero. Es fácil obtener las diferencias señaladas:

-0.4, -0.7, 0.6, 0.2, -0.7, -0.2, -0.4, -0.3

Colocando estos valores en orden numérico sin tener en cuenta el signo, tenemos:

-0.2, 0.2, -0.3, -0.4, -0.4, 0.6, -0.7, -0.7

Queda claro que la asignación de rangos a estos resultados presenta una dificultad, la de las *posiciones empatadas*. Hay dos resultados con el valor numérico 0.2, dos con un valor numérico 0.4, y dos con un valor numérico de 0.7, ¿cómo se calculan las posiciones? Este problema, en la práctica el único que uno se encuentra en los métodos de jerarquización, se resuelve asignando posiciones promedio a los valores empatados, con signos adecuados: por lo tanto, los rangos resultantes para estos datos son:

-1.5, 1.5, -3, -4.5, -4.5, -7.5, -7.5

En estos casos, merece la pena verificar si se ha hecho correctamente la ordenación calculando la suma de todos los valores, sin tener en cuenta el signo. La suma para los números anteriores es 36, la misma que la suma de los ocho primeros números naturales, y por tanto la ordenación es correcta. La suma de los valores positivos es 7.5, y la de los valores negativos es 28.5. El estadístico de la prueba es entonces 7.5. Al mirar la tabla A.12 se observa que, para  $n = 8$ , el estadístico de la prueba tiene que ser menor o igual que 3 para que se pueda rechazar la hipótesis nula al nivel de significación  $P = 0.05$ . En este caso, la hipótesis nula tiene que ser aceptada —no hay evidencia de la mediana (media) de la diferencia sea distinta de cero, y por tanto no hay ninguna evidencia sobre de una diferencia sistemática entre los dos métodos analíticos.

De estos ejemplos se desprende que la prueba de rangos y signos es un método simple y valioso. Su limitación principal radica en que no se puede aplicar a conjuntos de datos muy pequeños: para una prueba de dos colas al nivel de significación  $P = 0.05$ ,  $n$  tiene que ser al menos 6.

## 6.7 LA PRUEBA DE SUMA DE RANGOS DE WILCOXON Y MÉTODOS RELACIONADOS

La prueba de rangos y signos que se describe en la sección anterior es válida para el estudio de conjuntos individuales de mediciones y para conjuntos de datos por parejas que se puedan reducir con facilidad a conjuntos individuales. Sin embargo, en muchos casos es necesario comparar dos muestras independientes que no se pueda reducir a un conjunto único de datos. Dichas muestras pueden contener diferentes números de mediciones. Para abordar estos problemas se han propuesto varias pruebas no paramétricas. La más sencilla de comprender (aunque quizás no sea la más sencilla de realizar) es la prueba de suma de rangos de Wilcoxon. La ejecución de esta prueba se entiende mejor con un ejemplo.

*Ejemplo.* Se analizó una muestra de material fotográfico de desecho mediante espectrometría de absorción atómica para conocer el nivel de plata, proporcionando para cinco determinaciones sucesivas los siguientes valores: 9.8, 10.2, 10.7, 9.5 y 10.5  $\mu\text{g/ml}$ . Después de recibir un tratamiento químico se volvieron a analizar por el mismo procedimiento. Los resultados obtenidos para las cinco determinaciones sucesivas fueron los siguientes: 7.7, 9.7, 8.0, 9.9 y 9.0  $\mu\text{g/ml}$ . ¿Hay alguna evidencia de que el tratamiento produjo una reducción significativa en los niveles de plata?

El primer paso es ordenar *todos* los datos (es decir, de ambos conjuntos de mediciones). Para distinguir los resultados obtenidos después del tratamiento, éstos aparecen subrayados:

7.7, 8.0, 9.0, 9.5, 9.7, 9.8, 9.9, 10.2, 10.5, 10.7

Las posiciones —de nuevo subrayadas las que corresponden a las mediciones de la muestra tratada— son:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Queda claro que, incluso si aparecen posiciones empatadas y se tratan de la misma forma en la sección anterior, la suma de todas las posiciones tiene que ser  $n(n+1)/2$ , en este caso 55.

Si las mediciones de los dos conjuntos de resultados fueran indistinguibles, esperaríamos que tanto las posiciones subrayadas como las otras estuvieran mezcladas más o menos al azar en la lista. Esto es, la suma de las posiciones subrayadas y la de las otras posiciones debería ser similar (si el número de mediciones fuese el mismo en los dos conjuntos de resultados). En el presente caso la suma de las posiciones subrayadas (muestras tratadas) es 37. Estas sumas de posiciones se convierten ahora en los estadísticos  $T_1$  y  $T_2$  utilizando las ecuaciones:

$$T_1 = S_1 - n_1(n_1 + 1)/2 \quad (6.5)$$

$$T_2 = S_2 - n_2(n_2 + 1)/2 \quad (6.6)$$

donde  $S_1$  es la suma de las posiciones subrayadas y  $S_2$  es la otra suma (18 y 37 en este ejemplo) y  $n_1$  y  $n_2$  son los correspondientes números de observaciones. En este caso  $n_1 = n_2 = 5$ , con lo cual  $n_i(n_i + 1)/2 = 15$ , y  $T_1$  y  $T_2$  son 3 y 22 respectivamente. El menor de estos dos valores (3) se compara con el valor apropiado de la tabla A.13. Es especialmente importante observar que en este ejemplo podemos aplicar de forma ortodoxa una prueba de *una cola* porque el método de tratamiento está diseñado específicamente para *reducir* el nivel de plata del residuo fotográfico; jesto se podrá o no hacer con éxito, pero no hay ninguna posibilidad realista de que crezca realmente el nivel de plata! La tabla muestra que, para  $P = 0.05$  y  $n_1 = n_2 = 5$ , el menor valor de  $T$  no tiene que exceder de 4 si se va a rechazar la hipótesis nula. Así, es posible rechazar la hipótesis nula —que las dos muestras proceden de la misma población— y concluir que el tratamiento químico se ha mostrado efectivo en el momento de reducir el nivel de plata en el residuo fotográfico.

El funcionamiento de la prueba de suma de rangos se complica debido al cálculo de los valores de  $T$  a partir de las sumas de posiciones. Sin embargo, este paso adicional es

esencial ya que la prueba es también válida cuando  $n_1$  y  $n_2$  son distintos. En dichos casos no es razonable esperar que las sumas de los dos conjuntos de posiciones sean iguales, aunque se hayan mezclado más o menos al azar en la ordenación global, y las ecuaciones (6.5) y (6.6) tengan en cuenta este efecto. Debido a este cálculo adicional se han establecido dos pruebas alternativas, y probablemente más simples, para comparar dos conjuntos de datos que tienen diferente número de elementos.

La primera, la **prueba U de Mann-Whitney**, es un método más fácil e igualmente válido para calcular el menor valor de  $T$  definido antes.

*Ejemplo.* Aplique la prueba  $U$  de Mann-Whitney a los datos del ejemplo anterior.

En este procedimiento hay que encontrar el número de resultados en una muestra que supera cada uno de los valores en la otra muestra. En este ejemplo tenemos la impresión de que la concentración de plata de la solución tratada debería ser, en todo caso, menor que la solución no tratada. De esta forma esperamos encontrar que el número de casos en los que una muestra tratada tiene un valor más alto que una muestra no tratada debería ser pequeño. A continuación se hace una lista con cada uno de los valores de la muestra no tratada, y se anota el número de casos en que los valores para la muestra tratada son mayores.

Muestra no tratada	Valores más grandes en la muestra tratada	Número de valores más grandes
9.8	9.9	1
10.2	—	0
10.7	—	0
9.5	9.7, 9.9	2
10.5	—	0

La suma total de la tercera columna es 3, el mismo valor de  $T$  obtenido en la prueba de suma de rangos. Se puede demostrar que los dos métodos siempre producen el mismo resultado, con lo cual la tabla A.13 se utiliza tanto para la prueba  $U$  de Mann-Whitney como para la de suma de rangos. Es obvio que el método de la prueba  $U$  es una forma mucho más sencilla de llegar al valor de  $T$ . En realidad, cuando los números de mediciones son pequeños, el cálculo se puede hacer mentalmente, lo cual es una gran ventaja. Si se presentan empates (valores idénticos) en la prueba  $U$ , a cada uno se asigna un valor de 0.5 en el cálculo de  $T$ .

Otro método conveniente que tiene algunas características interesantes en la **prueba rápida de Tukey**. Parece apropiado que se ilustre su uso con el mismo ejemplo.

*Ejemplo.* Aplique la prueba rápida de Tukey a los datos del ejemplo anterior.

Se puede considerar que la prueba está formada por dos pasos, aun cuando sólo se disponga de pocos resultados; esos dos pasos se agrupan sin lugar a duda en un solo cálculo mental rápido. En el primero, se cuenta el número de resultados del segundo conjunto de datos que son *inferiores a todos los valores del primer conjunto*. Si no existe ningún valor con esa condición la prueba se da por terminada, y se acepta la hipótesis nula de medianas iguales. En el ejemplo que se está analizando, hay tres de esos valores, las lecturas 7.7, 8.0 y 9.0, son inferiores al menor valor del primer conjunto (9.5). En el segundo paso, se cuentan todos los valores del primer conjunto de datos que *son más grandes que todos los valores en el segundo conjunto*. De nuevo,

si no existen esos valores la prueba finaliza y se acepta la hipótesis nula. Aquí hay también tres valores, 10.2, 10.5 y 10.7, que superan el mayor valor del segundo conjunto de datos (9.7). (Esta aproximación se puede contrastar con la de la prueba *U* de Mann-Whitney, que identifica los valores *grandes* en la muestra que se espera tenga la mediana *más pequeña*.) Globalmente, por tanto, hay seis valores que no están dentro del intervalo donde se solapan las dos muestras. Este total (a menudo designado por *T*, pero que no se debe confundir con el estadístico *T* en la prueba de suma de rangos) es el estadístico de la prueba. El aspecto más interesante y destacado de la prueba rápida de Tukey es que normalmente no se necesitan tablas estadísticas para interpretar este resultado. Si suponemos que el número de lecturas de cada muestra no excede de 20, y que los dos tamaños muestrales no difieren mucho (condiciones que probablemente serían válidas en muchos experimentos analíticos realizados en laboratorios), los valores de referencia de *T* para un nivel de significación concreto serán *independientes del tamaño muestral*. Para una prueba de una cola se puede rechazar la hipótesis nula si  $T \geq 6$  (para  $P = 0.05$ ),  $\geq 7$  ( $P = 0.025$ ),  $\geq 10$  ( $P = 0.005$ ) y  $\geq 14$  ( $P = 0.0005$ ). (Para una prueba de dos colas los valores críticos de *T* a unos niveles de significación  $P = 0.05, 0.025, 0.005$  y  $0.0005$  son 7, 8, 11 y 15, respectivamente). En el ejemplo que estamos analizando, por tanto, el valor de *T* es suficientemente grande para ser significativo a un  $P = 0.05$  en una prueba de una cola. Podemos por lo tanto rechazar la hipótesis nula y concluir que el tratamiento reduce significativamente el contenido de plata en el material fotográfico de desecho, lo que concuerda con la prueba *U* de Mann-Whitney.

Si se producen empates en la prueba rápida de Tukey (es decir si uno de estos valores en la muestra hipotéticamente más grande es igual al mayor valor de la otra muestra, o si uno de los valores en la muestra “más pequeña” es igual al menor de la muestra “más grande”) entonces cada empate se cuenta como 0.5 en el cálculo del valor de *T*.

Además de su aplicación habitual para comparar medianas, la prueba de suma de rangos de Wilcoxon ha sido modificada por Siegel y Tukey para comparar la dispersión de dos conjuntos de resultados, y así ofrecer una alternativa verdaderamente no paramétrica a la prueba *F* (véase la Sec. 3.5). En esta versión de la prueba de suma de rangos, se juntan los datos y se disponen por orden creciente del valor numérico, como antes, restando uno de los conjuntos de resultados para diferenciarlo del otro. Luego se ordenan de una forma diferente: a la medición más pequeña se le asigna la posición 1, a la más grande se asigna la posición 2, la medición inmediatamente anterior a la más grande lleva la posición 3, la inmediatamente posterior a la más pequeña lleva la posición 4, la medición posterior a las dos más pequeñas lleva la posición 5 y así sucesivamente. (Si el número total de mediciones es impar se ignora la medida central.) Esta “ordenación alterna por pares” produce una situación en la que se asignan posiciones bajas a los resultados bajos y altos, y los resultados centrales reciben posiciones altas. Si un conjunto de datos tiene una dispersión significativamente mayor que el otro, la suma de sus posiciones debería ser mucho menor, mientras que si la dispersión es similar en los dos conjuntos de resultados, las sumas de sus posiciones también serán similares. La aplicación de este método en el ejemplo utilizado al principio de esta sección proporciona las posiciones siguientes:

Datos:	<u>7.7</u>	<u>8.0</u>	<u>9.0</u>	9.5	<u>9.7</u>	9.8	<u>9.9</u>	10.2	10.5	10.7
Posiciones:	1	4	5	8	9	10	7	6	3	2

La suma de las posiciones correspondientes a los valores subrayados (muestras que contienen plata tratada) es de 26, y la de las posiciones para las muestras no tratadas es de 29. Estas sumas se convierten entonces en los estadísticos de la prueba  $T_1$  y  $T_2$  como en las ecuaciones (6.5) y (6.6). El menor de estos valores, 11, el utilizado en la prueba, es mucho mayor que el valor crítico de 2 (para una prueba de *dos colas*, a un  $P = 0.05$ ) (véase la tabla A.13). La hipótesis nula, en este caso que la dispersión de los resultados es similar para los dos conjuntos de datos, debe ser aceptada. Si se analiza un poco esta prueba se verá que su validez queda reducida cuando los valores promedios para los dos conjuntos de datos son sustancialmente diferentes. En el caso extremo en que todas las mediciones de la muestra sean inferiores a todas las mediciones en la otra, los valores de  $T_1$  y  $T_2$  serán siempre tan parecidos como sea posible, cualquiera que sea la dispersión de las dos muestras. Si se teme que este efecto sea apreciable, se pueden estimar las medias de las dos muestras, y añadir la diferencia entre las medias a cada una de las mediciones del conjunto inferior. Esto eliminará cualquier efecto debido a medias diferentes, mientras que se mantendrá la dispersión de la muestra. Un ejemplo de aplicación a esta prueba se proporciona al final del capítulo.

## 6.8 PRUEBAS NO PARAMÉTRICAS SOBRE MÁS DE DOS MUESTRAS

En la sección anterior vimos pruebas que comparan dos muestras estadísticas entre sí. Sin embargo, los métodos no paramétricos no están limitados a dos conjuntos de datos: hay varios métodos disponibles que comparan tres o más muestras. Antes de dar una idea general sobre dos de estas pruebas, es importante mencionar un peligro que se debe evitar en todas las comparaciones de muchas muestras. Al examinar tres conjuntos de mediciones (por ejemplo), para comprobar si sus medianas son o no similares, existe la tentación de comparar únicamente las dos muestras que tengan las medianas mayor y menor. Esta aproximación simplista puede conducir a resultados engañosos. Cuando se toman varias muestras de la *misma población*, hay casos en que las medianas menor y mayor, consideradas en forma aislada, resultan ser significativamente diferentes. Esto se debe a que, a medida que crece el número de muestras, la diferencia entre las medianas menor y mayor tenderá a crecer. La aproximación correcta es realizar una prueba que considere *todas las muestras juntas*: si se demuestra que no todas podrían proceder de la misma población, entonces se pueden realizar pruebas separadas para intentar identificar dónde se producen las diferencias significativas. Aquí describimos, en términos generales, los principios de dos pruebas no paramétricas para tres o más conjuntos de datos; el lector que desee más información puede consultar los libros mencionados en la bibliografía.

La prueba de **Kruskal-Wallis** se puede considerar como una extensión del procedimiento de suma de rangos de Wilcoxon para comparar las medianas de tres o más muestras entre las que no se pueda considerar ningún tipo de emparejamiento. (Una generalización del análisis de plata descrito en la sección anterior, con tres muestras de material fotográfico de desecho, una sin tratamiento y las otras dos tratadas por métodos diferentes, proporcionaría un ejemplo donde sería útil la prueba.) Los resultados de las tres (o más) muestras se juntan y se disponen por orden de menor a mayor. Se determina el total de posiciones para los datos de las diferentes muestras: se promedia las posiciones con el mismo valor como en el método de Wilcoxon, aunque hay un procedimiento de corrección especial si se encuentran numerosos valores iguales. Es fácil apreciar que, si cada muestra tiene el mismo

número de mediciones (ésta no es una condición de la prueba), y si las muestras tienen medianas similares, entonces sería similar la suma total de posiciones en cada muestra, de manera que la suma de sus cuadrados sería mínima. De este modo, si tenemos tres muestras —cada una con cinco medidas— las posiciones irán de la 1 a la 15 y la suma de todas ellas será 120. Supongamos que las tres medianas son muy similares, y que la suma de las posiciones para cada muestra son por tanto iguales a 40. La suma de los cuadrados de estos totales será  $40^2 + 40^2 + 40^2 = 4800$ . Si las medianas son significativamente diferentes, entonces los totales de los rangos serán también diferentes unos de otros —digamos 20, 40 y 60. La suma de los cuadrados de dichos totales será siempre más grande que 4800 ( $20^2 + 40^2 + 60^2 = 5600$ ).

Se puede determinar la probabilidad de obtener cualquier suma de cuadrados utilizando el estadístico **chi cuadrada** (véase el Cap. 3). Si uno se refiere a las muestras por *A, B, C*, etc. (*k* muestras en total), con un número de medidas  $n_A, n_B, n_C$ , etc., y la suma de las posiciones por  $R_A, R_B, R_C$ , etc., entonces el valor de  $\chi^2$  está dado por:

$$\chi^2 = \frac{12}{N^2 + N} \left( \frac{R_A^2}{n_A} + \frac{R_B^2}{n_B} + \frac{R_C^2}{n_C} + \dots \right) - 3(N + 1) \tag{6.7}$$

donde  $N = n_A + n_B + n_C$  etc. Este valor  $\chi^2$  se compara con los valores usuales cuando el número total de mediciones es  $> ca. 15$ , no obstante se utilizan tablas especiales para números de mediciones más pequeños. El número de grados de libertad es  $k - 1$ . Los valores experimentales de  $\chi^2$  que superan los valores tabulados permiten rechazar la hipótesis nula (que las medianas de las muestras no son significativamente diferentes). Como ya se ha observado, en esta última situación se pueden realizar otras pruebas sobre pares individuales de muestras; de nuevo, el lector que desee mayor información puede consultar la bibliografía.

Ya hemos visto (Secs. 3.4 y 6.3) que cuando se comparan resultados *por parejas*, se pueden utilizar pruebas estadísticas especiales. Estas pruebas se basan en el siguiente principio: cuando dos métodos experimentales que no difieren significativamente se aplican a las mismas muestras químicas, las diferencias entre los pares de resultados emparejados deberían estar cerca de cero. Esto puede extenderse a tres o más conjuntos de resultados emparejados utilizando una prueba no paramétrica ideada en 1937 por el famoso economista americano, Milton Friedman. En química analítica, la principal aplicación de la **prueba de Friedman** es la comparación de tres (o más) métodos experimentales aplicados a las mismas muestras químicas. La prueba utiliza de nuevo el estadístico  $\chi^2$ , en este caso para calcular las diferencias que aparecen entre los valores de la suma total de posiciones, por distintos métodos. El siguiente ejemplo aclara la simplificación de la aproximación:

*Ejemplo.* Se determinaron los niveles de un pesticida en cuatro extractos de plantas por (A) cromatografía líquida de alta resolución, (B) cromatografía de gas-líquido, y (C) radioinmunoensayo. Se obtuvieron los siguientes resultados (todos en ng/ml):

Muestra	Método		
	A	B	C
1	4.7	5.8	5.7
2	7.7	7.7	8.5
3	9.0	9.9	9.5
4	2.3	2.0	2.9

¿Proporcionan los tres métodos valores para los niveles de pesticida que difieren significativamente?

Este problema se resuelve simplemente sustituyendo los resultados de la tabla por sus rangos. Para cada muestra se asigna la posición 1 al método que da el resultado mayor, tal como se indica a continuación:

Muestra	Método		
	A	B	C
1	1	3	2
2	1.5	1.5	3
3	1	3	2
4	2	1	3

Es necesario el uso de un valor promedio en el caso de posiciones empatadas en la muestra 2 (véase la Sec. 6.6). La suma de las posiciones para los tres métodos, A, B, C, son 5.5, 8.5 y 10, respectivamente. Estas sumas totalizarían  $nk(k + 1)/2$ , donde *k* es el número de métodos (aquí tres) y *n* el número de muestras (aquí cuatro). Se elevan al cuadrado las sumas de los rangos, lo que da 30.25, 72.25 y 100, respectivamente, y estos cuadrados se suman para dar el estadístico *R*, que en este caso es 202.5. El valor experimental de  $\chi^2$  se calcula entonces a partir de:

$$\chi^2 = \frac{12R}{nk(k + 1)} - 3n(k + 1) \tag{6.8}$$

que resulta en este caso 2.625. Al nivel de significación  $P = 0.05$  y con  $k = 3$ , los valores críticos de  $\chi^2$  son 6.0, 6.5, 6.4, 7.0, 7.1 y 6.2 para  $n = 3, 4, 5, 6, 7$  y 8, respectivamente. (Muchos conjuntos de tablas estadísticas proporcionan más datos, y cuando  $k > 7$ , se puede utilizar las tablas usuales de  $\chi^2$  con  $k - 1$  grados de libertad.) En este caso, el valor experimental de  $\chi^2$  es mucho menor que el valor crítico, y se debe aceptar la hipótesis nula: los tres métodos dan resultados que no difieren significativamente.

Es importante mencionar que la prueba de Friedman podría utilizarse alternativamente en forma inversa: suponiendo que los tres métodos analíticos dan resultados indistinguibles, se podría utilizar el mismo procedimiento para contrastar diferencias entre los cuatro extractos de plantas. En este caso *k* y *n* son 4 y 3 respectivamente; el lector puede verificar con cuidado que *R* es 270 y que el valor resultante de  $\chi^2$  es 9.0. Este es mayor que el valor crítico para  $P = 0.05, n = 3, k = 4$ , que es 7.4. Así, en esta segunda aplicación de la prueba se puede rechazar la hipótesis nula, y afirmar que las cuatro muestras *difieren* en sus niveles de pesticida. Se encuentran disponibles otras pruebas que permiten comparaciones entre pares de muestras seleccionados.

Evidentemente es mucho más simple realizar la prueba de Friedman en la práctica que el método ANOVA (véanse las Secs. 3.8, 3.10), aunque no tiene la capacidad de este último de estudiar los efectos de interacción (véase el Cap. 7).

### 6.9 CORRELACIÓN ORDINAL

Los métodos de jerarquización se pueden aplicar también a los problemas de correlación. El método de coeficiente de correlación ordinal de Spearman, que se describirá en esta

sección, es la aplicación más antigua de los métodos de ordenación en estadística, que data de 1904. Como otros métodos de ordenación, es especialmente ventajosa cuando uno de los dos conjuntos de observaciones en estudio puede expresarse *sólo* en términos de un orden de posición en lugar de expresarse en unidades cuantitativas. Así, en el siguiente ejemplo, se investiga la posible correlación entre las concentraciones de dióxido de azufre en un conjunto de vinos de mesa y la calidad de su sabor. La calidad del sabor de un vino no se expresa con facilidad en términos cuantitativos, pero para un panel de catadores de vinos es realmente sencillo ordenarlos según su preferencia. Ejemplos de otros atributos que se ordenan con facilidad pero que no se cuantifican fácilmente, incluyen la condición de animales experimentales, la calidad del espacio dedicado al laboratorio, y la eficiencia del personal del laboratorio. Se debería también recordar que si uno o los dos conjuntos de datos que se estudian fueran cuantitativos, entonces (en contraste con los métodos descritos en el capítulo 5) no es necesario que se distribuyan normalmente. Como otros estadísticos no paramétricos, el coeficiente de correlación ordinal de Spearman,  $\rho$ , es fácil de determinar e interpretar. Esto se muestra en el siguiente ejemplo.

*Ejemplo.* Un panel de expertos ordena según sus preferencias siete vinos de mesa diferentes; al mejor se le asigna la posición 1, al siguiente la 2, y así sucesivamente. El contenido de dióxido de azufre de cada vino (en partes por millón) se determina mediante análisis por inyección en flujo con detección colorimétrica. Utilice los resultados siguientes para determinar si hay relación entre la calidad percibida en el vino y el contenido de dióxido de azufre.

Vino	A	B	C	D	E	F	G
Rango según el sabor	1	2	3	4	5	6	7
Contenido de SO <sub>2</sub>	0.9	2.8	1.7	2.9	3.5	3.3	4.7

El primer paso en los cálculos es convertir las concentraciones de SO<sub>2</sub> de valores absolutos a valores de rango (si aparecen posiciones empatadas se promedian como se describió en las secciones anteriores):

Vino	A	B	C	D	E	F	G
Rango según el sabor	1	2	3	4	5	6	7
Contenido de SO <sub>2</sub>	1	3	2	4	6	5	7

Entonces, se calculan las diferencias,  $d_i$ , entre las dos ordenaciones. Claramente son 0, -1, 1, 0, -1, 1, 0. El coeficiente de correlación,  $\rho$ , está dado entonces por:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (6.9)$$

En este ejemplo, es  $1 - (24/336)$ , es decir, 0.929. La teoría demuestra que, como el coeficiente de correlación momento-producto, puede variar entre -1 y +1. Cuando  $n = 7$ , tiene que superar a 0.786 si se va a rechazar la hipótesis nula de ausencia de correlación al nivel de significación  $P = 0.05$  (véase la tabla A.14). Aquí, podemos concluir que hay una correlación entre el contenido de SO<sub>2</sub> de los vinos y su calidad percibida. Teniendo en cuenta la forma en que se definen las posiciones, ¡hay

una fuerte evidencia de que niveles de SO<sub>2</sub> más altos producen vinos menos aceptables al paladar!

Otro método de correlación ordinal, debido a Kendall, se introdujo en 1938; éste pretende tener algunas ventajas teóricas sobre el método de Sparman, pero es más complicado de calcular (especialmente cuando aparecen posiciones empatadas) y no se emplea con tanta frecuencia.

## 6.10 MÉTODOS DE REGRESIÓN NO PARAMÉTRICOS

Al analizar detalladamente los métodos de regresión lineales en el capítulo anterior, se hizo hincapié en la hipótesis de distribución normal para los errores en la dirección del eje  $y$ , y quedó clara la complejidad de alguno de los métodos de cálculo; ésta se supera en gran parte al utilizar calculadoras o computadores, y existen también algunos métodos de aproximación rápidos para ajustar líneas rectas a datos experimentales (véase la bibliografía). Se mantiene, por lo tanto, un cierto interés en las aproximaciones no paramétricas al problema de ajustar una línea recta a un conjunto de puntos: de los métodos disponibles, quizás el más simple es el **método incompleto de Theil**, denominado de esta forma para distinguirlo de otro procedimiento más complejo desarrollado por el mismo autor (el método "completo").

El método de Theil supone que un conjunto de puntos  $(x_1, y_1)$ ,  $(x_2, y_2)$ , etc., va a ser ajustado por una recta de la forma  $y = bx + a$ . El primer paso en los cálculos consiste en disponer los puntos en orden de  $x$  creciente. Si el número de puntos,  $x$ , es impar, el punto medio, es decir, el valor mediano de  $x$  se borra: el cálculo siempre exige un número par de puntos. Para cualquier par de puntos  $(x_i, y_i)$ ,  $(x_j, y_j)$  donde  $x_j > x_i$ , la pendiente,  $b_{ij}$ , de la línea que une los puntos se puede calcular a partir de:

$$b_{ij} = (y_j - y_i) / (x_j - x_i) \quad (6.10)$$

Se calculan las pendientes  $b_{ij}$  para el par de puntos  $(x_1, y_1)$  y el inmediatamente posterior al valor mediano de la  $x$ , para  $(x_2, y_2)$  y el segundo punto después del valor mediano de la  $x$ , y así sucesivamente hasta que se calcule la pendiente para la línea que une el punto inmediatamente anterior a la mediana de  $x$  con el último punto. De esta forma, si los datos originales contenían 11 puntos, se estimarían 5 pendientes (el punto mediano se ha omitido); 8 puntos originales proporcionarían 4 estimaciones de la pendiente, y así sucesivamente. Estas estimaciones de la pendiente, a su vez se disponen en orden ascendente y su valor mediano, calculado como se describió en la sección 6.2, es la pendiente estimada de la línea recta. Con este valor de  $b$ , los valores  $a_i$  de la ordenada en el origen se estiman para cada punto con la ayuda de la ecuación  $y = bx + a$ . De nuevo, las estimaciones de  $a$  se ordenan de forma ascendente y el valor mediano se elige como la mejor estimación de la ordenada en el origen de la recta. El método se aclara mediante el siguiente ejemplo.

*Ejemplo.* Se han obtenido los resultados siguientes en un experimento de calibración para la determinación absorciométrica de un complejo metálico tipo quelato:

Concentración, $\mu\text{g/ml}$ :	0	10	20	30	40	50	60	70
Absorbancia:	0.04	0.23	0.39	0.59	0.84	0.86	1.24	1.42

Utilice el método de Theil para estimar la pendiente y la ordenada en el origen de la recta que mejor se ajuste a esos puntos.

En este caso el cálculo se simplifica al tener un número par de observaciones, y por que los valores de  $x$  (es decir, las concentraciones) aparecen a intervalos regulares y están ya ordenadas. Por tanto se calculan las estimaciones de la pendiente de los cuatro pares de puntos:

$$b_{15} = (0.84 - 0.04)/40 = 0.0200$$

$$b_{26} = (0.86 - 0.23)/40 = 0.0158$$

$$b_{37} = (1.24 - 0.39)/40 = 0.0212$$

$$b_{48} = (1.42 - 0.59)/40 = 0.0208$$

Ahora se ordenan estas estimaciones de la pendiente, resultando 0.0158, 0.0200, 0.0208, 0.0212. La estimación mediana de la pendiente es, por lo tanto, el promedio entre 0.0200 y 0.0208, es decir, 0.0204. Se utiliza este valor de  $b$  para estimar la ordenada en el origen,  $a$ ; los ocho valores  $a_i$  individuales son:

$$a_1 = 0.04 - (0.0204 \times 0) = +0.040$$

$$a_2 = 0.23 - (0.0204 \times 10) = +0.026$$

$$a_3 = 0.39 - (0.0204 \times 20) = -0.018$$

$$a_4 = 0.59 - (0.0204 \times 30) = -0.022$$

$$a_5 = 0.84 - (0.0204 \times 40) = +0.024$$

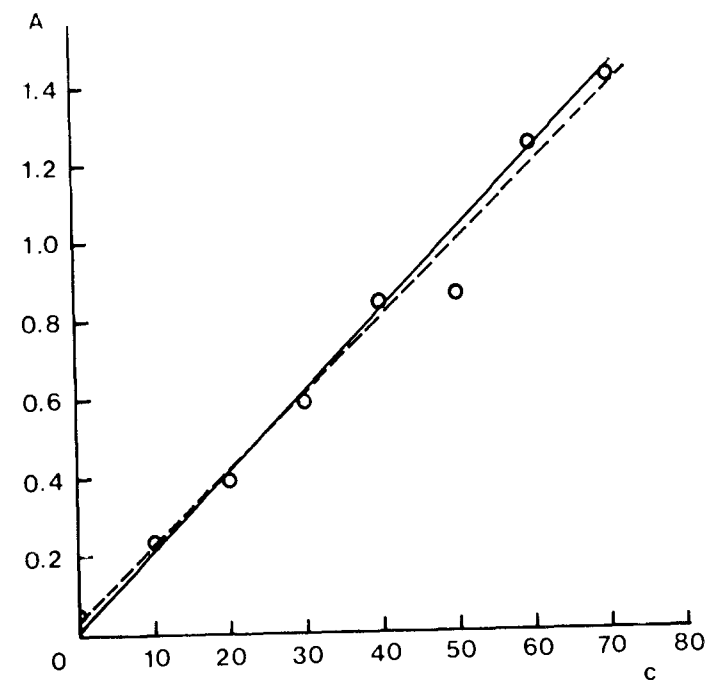
$$a_6 = 0.86 - (0.0204 \times 50) = -0.160$$

$$a_7 = 1.24 - (0.0204 \times 60) = +0.016$$

$$a_8 = 1.42 - (0.0204 \times 70) = +0.008$$

Ordenando estas estimaciones de la ordenada en el origen, tenemos -0.160, -0.022, -0.018, -0.008, +0.016, +0.024, +0.026, +0.040. La estimación mediana es, por lo tanto, +0.004. De este modo concluimos que la mejor línea recta está dada por  $y = 0.0204x + 0.004$ . Es fácil mostrar que la recta de "mínimos cuadrados", calculada por los métodos del capítulo 5, es  $y = 0.0195x + 0.019$ . La figura 6.1 muestra que ambas rectas son similares cuando se representan gráficamente. Sin embargo, el método de Theil tiene tres ventajas distintas: no supone que todos los errores están en la dirección de  $y$ ; no supone que los errores en la dirección de  $x$  o de  $y$  se distribuyan normalmente, y no se ve afectado por la presencia de resultados anómalos. Este último aspecto queda claramente ilustrado por el punto (50, 0.86) de este ejemplo; tiene toda la apariencia de ser un resultado anómalo, pero su valor no afecta, en absoluto, el cálculo de Theil, ya que ni  $b_{26}$  ni  $a_6$  afectan directamente las estimaciones medianas de la pendiente y ordenada en el origen, respectivamente. En el cálculo de mínimos cuadrados, sin embargo, este resultado anómalo tiene tanto peso como los otros; esto se refleja en los resultados calculados: la línea de mínimos cuadrados pasa más cerca del resultado anómalo que la línea no paramétrica. Se dice que los métodos estadísticos que no se ven muy afectados por los resultados anómalos son métodos "robustos": su importancia en éste y otros casos está clara.

A diferencia de la mayoría de los métodos no paramétricos, el método de Theil conlleva cálculos tediosos, y sería de gran utilidad un programa de computador. Tenga en cuenta



**Figura 6.1** Gráfica de la recta de calibración, calculada por el método de Theil (—) y por el método de mínimos cuadrados del capítulo 5 (- - -).

que también existen métodos no paramétricos disponibles para ajustar curvas, si bien éstos escapan a los objetivos de este libro.

## 6.11 LA PRUEBA DE BONDAD DEL AJUSTE DE KOLMOGOROV

En el capítulo 3 se estudió el problema estadístico frecuente de la "bondad del ajuste". Éste surge cuando se necesita probar la procedencia de una muestra de observaciones de una distribución concreta, como puede ser una distribución normal. La prueba de chi cuadrada es muy adecuada para este propósito cuando los datos se presentan como frecuencias, aunque la prueba no se utiliza normalmente para menos de 50 observaciones, y es difícil usarla con datos continuos. En esta sección se describe el método de Kolmogorov, que es muy apropiado para probar la bondad del ajuste con datos continuos y aislados. Este método se puede extender a la comparación de dos muestras, aunque no se describirá con detalle. Estos métodos modificados fueron descubiertos en primer lugar por Smirnov, y al conjunto de estas pruebas se les conoce, a menudo, como el método de Kolmogorov-Smirnov.

El principio de la aproximación de Kolmogorov es muy simple. Compara la curva de frecuencias acumuladas de los datos que se desea contrastar con la curva de frecuencias acumuladas de la distribución propuesta como hipótesis. El concepto de curva de frecuencias acumuladas, y su aplicación con relación al papel de probabilidad normal, se analizó en el capítulo 3. Cuando se han trazado las curvas experimental e hipotética, el estadístico de la prueba se obtiene encontrando la máxima diferencia vertical entre ambas, y compa-



rando este valor, de la forma habitual, con el conjunto de valores tabulados. Si los datos experimentales se apartan sustancialmente de la distribución esperada, se esperará que las dos curvas se encuentren separadas en parte del diagrama de frecuencias acumuladas; sin embargo, si los datos están en gran consonancia con la distribución esperada, las dos curvas nunca se separarán mucho una de otra. En la práctica, el método de Kolmogorov tiene dos aplicaciones frecuentes —prueba de aleatoriedad, y prueba de normalidad de una distribución—; el funcionamiento del procedimiento se ilustrará con ejemplos simples.

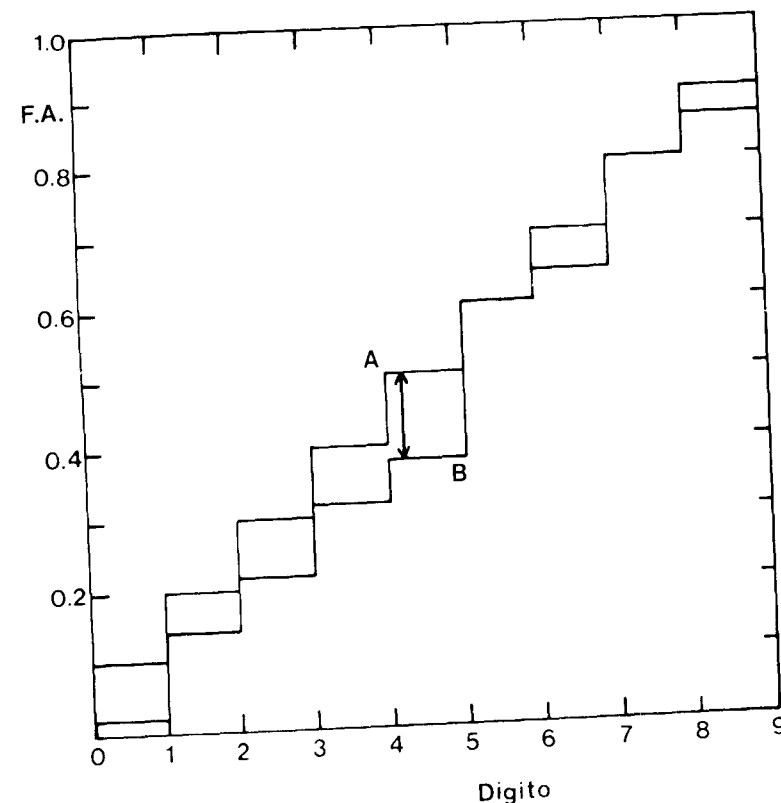
*Ejemplo.* Se analizaron las lecturas de bureta obtenidas por un laboratorio durante un periodo dilatado. En 50 lecturas se ha encontrado que los últimos dígitos registrados son los que aparecen en la siguiente tabla:

Dígito:	0	1	2	3	4	5	6	7	8	9
Frecuencia:	1	6	4	5	3	11	2	8	3	7

¿Son estas lecturas congruentes con un uso aleatorio de los dígitos 0-9? (Note que los datos de este ejemplo son idénticos a los del ejercicio 5 al final del capítulo 3.)

En este ejemplo la distribución hipotética es inmediata: se presupone que todos los dígitos del 0 al 9 tienen la misma probabilidad de aparecer, y no son posibles otros resultados. La curva hipotética de frecuencias acumuladas es, por lo tanto, una función regular escalonada como se muestra en la figura 6.2. Las frecuencias acumuladas en la práctica son claramente: 0.02, 0.14, 0.22, 0.32, 0.38, 0.60, 0.64, 0.80, 0.86, 1.00 en  $x = 0, 1, 2, \dots, 9$ , respectivamente. Las dos curvas muestran una diferencia máxima de 0.12 en  $x = 4$ . (En ejemplos simples de este tipo es innecesario representar en la práctica las frecuencias acumuladas; la diferencia máxima se puede obtener examinando los datos.) La tabla A.15, que es adecuada para una prueba de dos colas (el caso usual), muestra que al nivel de significación  $P = 0.05$  para  $n = 50$  el valor crítico es 0.188, de manera que no se puede rechazar la hipótesis nula de que los dígitos se distribuyan aleatoriamente.

El lector observará que este resultado es el mismo que el obtenido utilizando el método  $\chi^2$  cuadrada (véase el Cap. 3, ej. 5). Una inspección minuciosa de las dos pruebas muestra que, en el cálculo de  $\chi^2$  cuadrada, los valores experimental y tabulado son casi los mismos, es decir, los dígitos de titulación no se desvían significativamente de una distribución aleatoria. En la prueba de Kolmogorov, sin embargo, los resultados sugieren que los datos se desvían ostensiblemente de la aleatoriedad. Esta reflexión surge del hecho de que los valores de la tabla A.15 están basados en la hipótesis de que la distribución probada es continua. Cuando los datos son discretos, los resultados de la prueba son conservadores, es decir, tienden a dar muy pocos resultados significativos. No obstante hay algunas ventajas en la prueba de Kolmogorov. En los ejemplos del tipo que se acaba de analizar, es más sencillo de utilizar, y es más fiable cuando  $n$  es pequeño. (El ejemplo anterior utiliza precisamente el conjunto de datos más pequeños que generalmente se puede estudiar por el método  $\chi^2$  cuadrada: 50 observaciones y una frecuencia esperada de 5 para cada número.) El método de Kolmogorov también proporciona límites de confianza no paramétricos para la función de distribución acumulada verdadera, utilizando los datos de la tabla A.15. Este tema se trata con mayor profundidad en pruebas detalladas sobre métodos no paramétricos (véase la bibliografía).



**Figura 6.2** El empleo del método de Kolmogorov para probar la aleatoriedad de los dígitos 0-9. La diferencia máxima entre la representación gráfica de la frecuencia acumulada teórica (A) y la representación experimental (B), se muestra con las flechas ( $\longleftrightarrow$ ).

Cuando se utiliza el método de Kolmogorov para probar si una distribución es normal, lo primero que hay que hacer es transformar los datos originales, que podrían tomar cualquier valor para la media y la desviación estándar, en la **variable normal estandar**,  $z$ . Esto se hace utilizando la ecuación:

$$z = (x - \mu) / \sigma \quad (6.11)$$

en la que cada término tiene su significado habitual. Se suele considerar esta transformación como la conversión de datos en *forma tipificada o estandar*. La teoría demuestra que  $z$ , calculado en esta forma, se distribuye normalmente, y su función de distribución está dada en muchas colecciones de tablas estadísticas. La ecuación (6.11) se puede utilizar de dos formas. A veces es necesario probar la procedencia de un conjunto de datos de una distribución normal *concreta*, de media y desviación estándar *dadas*. En tal caso, se transforman directamente los datos experimentales mediante la ecuación (6.11) y se realiza la prueba de Kolmogorov. Con más frecuencia, será necesario contrastar simplemente la procedencia de los datos de *cualquier* distribución normal. En este caso primero se estima la media y la desviación estándar, por los métodos simples del capítulo 2; a continuación

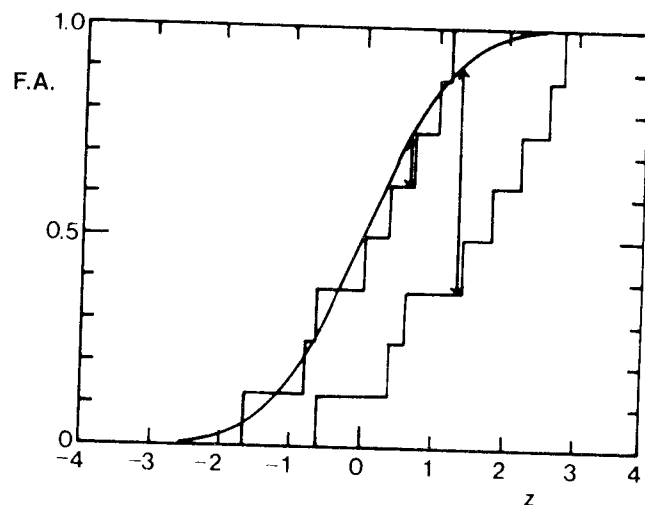


se transforman por la ecuación (6.11); y por último se aplica el método Kolmogorov. Estos dos tipos de pruebas se analizan en el ejemplo siguiente.

*Ejemplo.* Se realizaron ocho titulaciones, con los resultados 25.13, 25.02, 25.11, 25.07, 25.03, 24.97, 25.14 y 25.09 ml. ¿Dichos resultados podrían proceder de (a) una población normal con media 25.00 ml y desviación estándar 0.05 ml, y (b) de cualquier otra población normal?

(a) En este caso, el primer paso es transformar los valores  $x$  en valores  $z$ , utilizando la relación  $z = (x - 25.00)/0.05$  obtenida de la ecuación (6.11). Los ocho resultados son transformados en 2.6, 0.4, 2.2, 1.4, 0.6, -0.6, 2.8 y 1.8. Estos valores  $z$  se ordenan y representan gráficamente como una función de distribución con un salto de 0.125 (es decir,  $1/8$ ) en cada paso. (Observe que éste no es el mismo cálculo que se utilizó en la sección 3.12.) La comparación con la función hipotética para  $z$  indica (véase la Fig. 6.3) que la diferencia máxima es 0.545 en un valor  $z$  ligeramente inferior a 1.4. Para probar este valor se utiliza la tabla A.16; la tabla A.15 no es adecuada después que los datos se hayan transformado en la forma estándar. Según la tabla A.16, para  $n = 8$  y  $P = 0.05$ , el valor crítico es 0.288, de modo que se puede rechazar la hipótesis nula —las titulaciones probablemente no proceden de una población normal con media 25.00 ml y desviación estándar 0.05 ml.

(b) En este caso, se estiman la media y la desviación estándar [utilizando las Ecs. (2.1) y (2.2)] por 25.07 y 0.059 ml, respectivamente, y el último resultado tiene dos cifras significativas correctas. Los valores de  $z$  están ahora dados por  $z = (x - 25.07)/0.059$ , es decir, por 1.02, -0.85, 0.68, 0, -0.68, -1.69, 1.19, 0.34. El diagrama de frecuencias acumuladas para estos valores difiere de la curva hipotética a lo sumo por 0.125 (en varios puntos). Esta diferencia es mucho más pequeña que el valor crítico de 0.288. De esta forma podemos aceptar la hipótesis nula de que los datos proceden de una población normal con una media 25.07 y una desviación estándar 0.059.



**Figura 6.3** El empleo del método de Kolmogorov para probar la distribución normal. Las diferencias máximas entre la curva de frecuencias acumuladas teórica y las dos distribuciones probadas se reflejan con las flechas ( $\longleftrightarrow$ ).

## 6.12 CONCLUSIONES

Las pruebas de ejecución rápida y no paramétricas, descritas en este capítulo son sólo una pequeña parte de todos los métodos de este tipo que hay disponibles. Los ejemplos expuestos manifiestan claramente tanto su potencia como sus limitaciones. En muchos casos su rapidez y conveniencia aventajan, en cierta medida, a los métodos convencionales, y las pruebas no paramétricas no conllevan la hipótesis de una distribución normal. Se ajustan perfectamente al examen preliminar de un pequeño número de mediciones, y a cálculos hechos con rapidez —a menudo sin la necesidad de tablas— mientras el analista está en el banco de pruebas o en la fábrica. Se puede utilizar también cuando se estudian tres o más muestras (Sec. 6.8). La **potencia** (es decir, la probabilidad de que una hipótesis nula falsa sea rechazada: véase la Sec. 3.13) de un método no paramétrico puede ser menor que la de la correspondiente prueba paramétrica, pero la diferencia se da sólo en raras ocasiones. Por ejemplo, se han hecho muchas comparaciones de las potencias en condiciones diferentes (es decir, distribuciones poblacionales y tamaños muestrales diferentes) de la prueba de rangos de Wilcoxon y de la prueba  $t$ . La primera opera muy bien en casi todas las circunstancias y sólo es marginalmente menos potente que la prueba  $t$ , incluso cuando los datos proceden de una población normalmente distribuida. Muchos programas para computadores personales incluyen ahora varias pruebas no paramétricas. Dichos programas permiten evaluar un conjunto concreto de datos con rapidez por dos o más métodos, y parece seguro que aumenta el interés en estos métodos en lugar de reducirse, ya que resultan muy útiles.

## BIBLIOGRAFÍA

Conover, W. J., *Practical Non-parametric Statistics*, Nueva York, Wiley, 1971. Probablemente el mejor texto general de métodos no paramétricos.

Daniel, W. W., *Applied Nonparametric Statistics*, Boston, Houghton Mifflin, 1978. Es un texto muy general que abarca un amplio espectro de métodos no paramétricos con bastante profundidad; contiene muchos ejemplos.

Langley, R., *Practical Statistics*, Londres, Pan Books, 1968. Es un texto bastante asequible, con especial énfasis en las pruebas no paramétricas.

Sprent, P., *Quick Statistics*, Londres, Penguin Books, 1981. Es una introducción excelente a los métodos no paramétricos: claro, explicaciones con pocas matemáticas, y muchos ejemplos y ejercicios.

## EJERCICIOS

- Se realizó una titulación 4 veces, con los siguientes resultados: 9.84, 9.91, 9.89, 10.20 ml. Calcule y comente la mediana y la media de estos resultados.
- El nivel de azufre en lotes de combustible de un avión, según manifiesta el fabricante, está distribuido simétricamente con un valor mediano del 0.1%. Se ha encontrado que lotes sucesivos tienen concentraciones de azufre de 0.09, 0.12, 0.10, 0.11, 0.08, 0.17, 0.12, 0.14 y 0.11%. Utilice la prueba de signos, la prueba de rangos y signos, y una prueba de recorridos para probar la afirmación del fabricante.

3. La concentración (g/100 ml) de inmunoglobulina G en suero sanguíneo de 10 donantes se mide por inmunodifusión radial (i.d.r) y por electroinmunodifusión (e.i.d), con los siguientes resultados.

Donantes:	1	2	3	4	5	6	7	8	9	10
Resul. i.r.d.	1.3	1.5	0.7	0.9	1.0	1.1	0.8	1.8	0.4	1.3
Resul. e.i.d.	1.1	1.6	0.5	0.8	0.8	1.0	0.7	1.4	0.4	0.9

¿Son significativamente diferentes los resultados de los dos métodos?

4. En el desarrollo de un nuevo método para la determinación de niveles de alcohol en la sangre, se analizó 5 veces una única muestra de sangre, con los resultados 64.5, 66.0, 63.9, 65.1 y 64.0 mg/100 ml. El método de análisis estándar, aplicado a la misma muestra, proporciona los resultados 66.2, 65.8, 66.3 y 65.6 mg/100 ml. Utilice pruebas estadísticas basadas en los recorridos de una serie de resultados para contrastar si (a) los dos métodos proporcionan valores medios significativamente diferentes, y (b) si difieren significativamente en la precisión.
5. Se encontró que diez barras de grafito utilizadas sucesivamente en un espectrofotómetro de absorción atómica electrotermico duraron 24, 26, 30, 21, 17, 23, 22, 25 y 25 unidades de tiempo. Pruebe la aleatoriedad de los tiempos de funcionamiento de estas barras.
6. Los niveles de alcohol en sangre encontrados para cinco voluntarios, después de beber tres jarras de cerveza cada uno, fueron de 104, 79, 88, 120 y 90 mg/100 ml. Otro conjunto de voluntarios bebieron tres jarras de cerveza rubia cada uno, y se encontró que los niveles de alcohol en sangre eran de 68, 86, 71, 79, 91 y 66 mg/100 ml. Utilice la prueba de suma de intervalos, la prueba rápida de Tukey o la prueba de Mann-Whitney para investigar si es cierto que al beber cerbeza rubia se produce un nivel de alcohol en la sangre inferior que al beber la misma cantidad de otro tipo de cerveza. Utilice el método de Siegel-Tukey para verificar si las dispersiones en estos dos conjuntos de resultados difieren significativamente.
7. Un laboratorio universitario de química posee siete espectrofotómetros de absorción atómica (A-G). Las encuestas sobre la opinión de los estudiantes de investigación y del personal académico de plantilla demuestran que el orden de preferencia de los estudiantes con relación a estos instrumentos es B, G, A, D, C, F, E y que el del personal académico de plantilla es G, D, B, E, A, C, F. ¿Se encuentran correlacionadas las opiniones de los estudiantes y del personal académico?
8. Utilice el método de Thiel para calcular la línea de regresión para los datos del ejercicio 1 del capítulo 5.
9. Tomando los datos de inmunodifusión radial del ejercicio 3, emplear el método Kolmogorov para probar la hipótesis de que los niveles de inmunoglobulina G del suero se distribuyen normalmente con una media de 1.0 g/100 ml y una desviación estándar de 0.2 g/100 ml. ¿Se ajustará mejor cualquier otra distribución normal a los datos?
10. Los niveles de níquel en tres muestras de aceite crudo se determinaron por espectrometría de absorción atómica, con los siguientes resultados:

Muestra	Medias (Ni, ppm)					
1	14.2	16.8	15.9	19.1	15.5	16.0
2	14.5	20.0	17.7	18.0	15.4	16.1
3	18.3	20.1	16.9	17.7	17.9	19.3

Utilice el método de Kruskal-Wallis para decidir si los niveles de níquel en los tres aceites difieren significativamente.

## Diseño experimental, optimización y reconocimiento de pautas

### 7.1 INTRODUCCIÓN

En el capítulo 3 se introdujo la idea de un factor. Un factor es cualquier aspecto de las condiciones experimentales que influye en el resultado obtenido en un experimento. La sección 3.9 mostró el ejemplo de la dependencia existente entre una señal de fluorescencia y las condiciones en las cuales se almacenó una solución. En este experimento el factor de interés eran las condiciones de almacenamiento; en este caso se le asignó el nombre de factor controlado ya que el experimentador podía alterarlo a su voluntad. En otro ejemplo del capítulo siguiente (Sec. 4.3), en el que se probaba la pureza de la sal de diferentes partes de un barril, el factor de interés, es decir, la parte del barril de la que se tomaba la sal, se elegía al azar; por esta razón se denomina factor incontrolado. Estos dos factores son **cualitativos** ya que sus posibles "valores" no se pueden ordenar numéricamente. Un factor para el cual los valores posibles se *pueden* ordenar numéricamente se denomina **cuantitativo**, por ejemplo, la temperatura. Se entiende por niveles diferentes de un factor los valores diferentes que dicho factor toma.

En los capítulos 3 y 4 se presentaron estos experimentos como una introducción a los cálculos que aparecen en el análisis de la varianza (ANOVA). No se hizo mención alguna de las condiciones experimentales que podrían haber introducido otros factores que afectasen los resultados. En el experimento de fluorescencia dichos posibles factores incluyen: la temperatura ambiente, el fluorímetro utilizado para cada medición, la persona que realiza las mediciones y el día de realización del experimento. Cualquiera de estos factores *podría* haber influido en los resultados para dar el comportamiento observado e invalidar la conclusión extraída sobre el efecto de las condiciones de almacenamiento. Queda claro que si se quiere que las conclusiones extraídas de un experimento sean correctas, hay que identificar los diferentes factores que influyen en el resultado y, si es posible, controlarlos. El término **diseño experimental** se utiliza habitualmente para describir las etapas de (1) identificación de aquellos factores que pueden influir en el resultado de un experimento, (2) diseño del experimento de modo que se minimicen los

efectos de los factores incontrolados, y (3) utilización del análisis estadístico para separar los efectos de los diferentes factores implicados.

## 7.2 ALEATORIZACIÓN

Una de las hipótesis del ANOVA es que la variación incontrolada es aleatoria. No obstante, en aquellas mediciones tomadas durante un periodo, la variación en un factor incontrolado como la presión, la temperatura, el deterioro del aparato, etc., puede producir una tendencia en los resultados. Como consecuencia, los errores debidos a la variación incontrolada no son aleatorios durante mucho tiempo, ya que los errores sucesivos están correlacionados. Esto puede conducir a un error sistemático en los resultados. Afortunadamente este problema se supera utilizando la técnica de **aleatorización**. Supongamos que deseamos comparar el efecto de un único factor con tres niveles diferentes. En un experimento de este tipo los niveles diferentes se llaman a menudo **tratamientos**. Este término tiene su origen en el desarrollo que hace Fisher del ANOVA al analizar experimentos propios de la agricultura. En estos casos, los diferentes tratamientos podrían ser, por ejemplo, la aplicación de fertilizantes distintos. Notemos los tres tratamientos diferentes A, B, C y exijamos que se realicen cuatro mediciones repetidas en cada uno de los tratamientos. En lugar de tomar las cuatro mediciones para el tratamiento A, seguidas de las otras cuatro para B, y luego de las otras cuatro para C, podemos llevar a cabo doce mediciones en un orden aleatorio que se decide mediante una tabla de números aleatorios. Para ello se asigna a cada tratamiento un número para cada repetición:

A	B	C
01 02 03 04	05 06 07 08	09 10 11 12

(Observe que cada número tiene la misma cantidad de dígitos.) Después entramos en una tabla de números aleatorios (véase la tabla A.6) en un punto arbitrario y leemos los pares de dígitos, descontando los pares de la forma 00, 13-99, y se repite el proceso doce veces. Supongamos que como resultado de lo anterior aparece la sucesión 02, 10, 04, 03, 11, 01, 12, 06, 08, 07, 09, 05, entonces se realizarían las mediciones en los niveles diferentes en el orden A, C, A, A, C, A, B, B, B, C, B. Este orden aleatorio de mediciones asegura que los errores causados en el nivel A por factores incontrolados sean aleatorios, así como en los niveles B y C.

## 7.3 EXPERIMENTOS CON BLOQUES

Un inconveniente de la aleatorización completa es que no aprovecha las posibles subdivisiones naturales en el material experimental. Suponga, por ejemplo, que no se pudiesen llevar a cabo las doce mediciones del ejemplo anterior en el mismo día y se tuvieran que obtener en cuatro días consecutivos. Utilizando el mismo orden de antes resultaría:

Día 1	A	C	A
Día 2	A	C	A
Día 3	C	B	B
Día 4	B	C	B

Con este diseño, las mediciones que utilizan el tratamiento A se llevan a cabo en los primeros días, mientras que aquellas que utilizan el tratamiento B se realizan en los últimos dos días. Si se tuviese la impresión de que existe diferencia entre los tratamientos A y B no sería posible decir si esta diferencia era auténtica o si fue provocada por el efecto de utilizar los dos tratamientos en pares de días diferentes. Un mejor diseño es aquel en el que cada tratamiento se utiliza una vez cada día, y el orden de los tratamientos se aleatoriza cada día. Por ejemplo:

Día 1	A	C	B
Día 2	A	B	C
Día 3	C	B	A
Día 4	C	A	B

Un grupo que contenga una medición para cada tratamiento recibe el nombre de **bloque** (de nuevo un término de origen agrícola que significa una zona de un campo) y el diseño se llama un diseño de **bloques aleatorizados**. En este caso, el conjunto de las mediciones de cada día constituye un bloque.

Otra ventaja de un experimento por bloques es que es más sensible que otro sin bloques, como un método de probar si tratamientos diferentes producen resultados significativamente diferentes. El ANOVA —en este caso de dos factores ya que cada medición se clasifica de acuerdo con dos factores, tratamiento y día— permite la separación de las tres fuentes de variación: variación entre bloques, variación entre tratamientos y variación aleatoria debida al error experimental. Para decir si los tratamientos producen resultados significativamente diferentes, se compara la estimación de la varianza entre tratamientos con la varianza estimada de los errores aleatorios. La sensibilidad del experimento depende del tamaño de la variación aleatoria: cuando más pequeña sea la variación aleatoria, más pequeña será también la diferencia que se pueda detectar entre los tratamientos. En un experimento sin bloques, la variación aleatoria sería más grande, ya que incluiría una contribución debida a la variación día a día y, en consecuencia, se reduciría la sensibilidad del experimento.

## 7.4 ANOVA DE DOS FACTORES

En un ANOVA de dos factores, cada medición,  $x_{ij}$ , se clasifica con dos factores, como se muestra en la tabla 7.1. Hay  $N$  mediciones divididas entre  $c$  niveles de tratamiento y  $r$  bloques (por tanto  $N = cr$ ). También aparecen los totales por columna y por fila y el gran total,  $T$ , necesarios en los cálculos. La deducción de las fórmulas empleadas no se dará con detalle como se hizo para el ANOVA de un factor en la sección 3.9. El principio del método es el mismo, y al final del capítulo se proporcionan referencias para los que estén interesados en los aspectos matemáticos del tema. Las fórmulas para el cálculo de la variación de las tres fuentes diferentes, o sea, entre tratamientos, entre bloques y error experimental, se dan en la tabla 7.2.

*Ejemplo.* En un experimento para comparar el porcentaje de eficiencia en la extracción de iones metálicos en solución acuosa por agentes quelantes, se obtuvieron los siguientes resultados:

**Tabla 7.1** Forma general de la tabla del ANOVA de dos factores

	Tratamiento						
	1	2	.....	$j$	.....	$c$	Totales de fila
Bloque 1	$x_{11}$	$x_{12}$	.....	$x_{1j}$	.....	$x_{1c}$	$T_{1.}$
Bloque 2	$x_{21}$	$x_{22}$	.....	$x_{2j}$	.....	$x_{2c}$	$T_{2.}$
.....	.....	.....	.....	.....	.....	.....	.....
Bloque $i$	$x_{i1}$	$x_{i2}$	.....	$x_{ij}$	.....	$x_{ic}$	$T_{i.}$
.....	.....	.....	.....	.....	.....	.....	.....
Bloque $r$	$x_{r1}$	$x_{r2}$	.....	$x_{rj}$	.....	$x_{rc}$	$T_{r.}$
Totales de columna	$T_{.1}$	$T_{.2}$	.....	$T_{.j}$	.....	$T_{.c}$	$T = \text{Gran total}$

Días	Agente quelante			
	A	B	C	D
1	84	80	83	79
2	79	77	80	79
3	83	78	80	78

Cada día se prepararon disoluciones del ion metálico (con concentración especificada) y se realizó la extracción con cada uno de los agentes quelantes, tomados de una manera aleatoria.

En este experimento el uso de diferentes agentes quelantes es un factor controlado, ya que el investigador los elige; el día de la ejecución del experimento supone un factor incontrolado ya que introduce variación incontrolada causada, a la vez, por los cambios en la temperatura del laboratorio, presión, etc., y por ligeras diferencias en la concentración de la disolución del ion metálico, es decir, el día es un factor aleatorio. En los capítulos 3 y 4 se explicó la manera de utilizar el ANOVA para probar un efecto significativo debido a un factor controlado o para estimar la varianza de un factor incontrolado. En este experimento —en que hay a la vez dos tipos de factores— el ANOVA de dos factores se puede utilizar de dos maneras: (i) para probar si los diferentes agentes quelantes tienen eficiencias significativamente diferentes, y (ii) para

**Tabla 7.2** Fórmulas para ANOVA de dos factores

Fuente de variación	Suma de cuadrados	Grados de libertad
Entre tratamientos	$\sum_{j=1}^c T_{.j}^2 / r - T^2 / N$	$c - 1$
Entre bloques	$\sum_{i=1}^r T_{i.}^2 / c - T^2 / N$	$r - 1$
Residual	por diferencia	por diferencia
Total	$\sum_i \sum_j x_{ij}^2 - T^2 / N$	$N - 1$

probar si la variación día a día es significativamente más grande que la variación debida al error aleatorio de la medición y, si es así, estimar su varianza.

Como en un ANOVA de un factor, los cálculos se pueden simplificar restando a cada medición un número arbitrario. En la siguiente tabla se muestran mediciones una vez que se les ha restado 80 a cada uno de ellas.

Bloques	Tratamientos				Totales por filas, $T_i$	$T_i^2$
	A	B	C	D		
Día 1	4	0	3	-1	6	36
Día 2	-1	3	0	-1	5	25
Día 3	3	-2	0	-2	-1	1
Totales por columna, $T_j$	6	-5	3	-4	0 = Gran total, $T$	
$T_j^2$	36	25	9	16		
$\sum_j T_j^2 = 86$						

Tenemos:  $r = 3$ ,  $c = 4$ ,  $N = 12$ ,  $\sum_i T_i^2 = 62$ ,  $\sum_j T_j^2 = 86$ ,  $T = 0$ ,  $\sum_i \sum_j x_{ij}^2 = 54$ .

En la siguiente tabla se muestra el cálculo de los cuadrados medios para cada fuente de variación.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio
Entre tratamientos	$86/3 - 0^2/12 = 28.6667$	3	$28.6667/3 = 9.5556$
Entre bloques	$62/4 - 0^2/12 = 15.5$	2	$15.5/2 = 7.75$
Residual	por diferencias = 9.8333	6	$9.8333/6 = 1.6389$
Total	$54 - 0^2/12 = 54.0$	11	

(Ya que para obtener la suma de cuadrados residual hay que calcular una diferencia, se deben tomar tantas cifras decimales como sean posibles para evitar errores en esta cantidad.)

Para el lector puede ser formativo el verificar que las sumas de cuadrados *separan* los efectos de bloque y tratamiento. Esto puede hacerse, digamos, incrementando todos los valores de un bloque en una cantidad fija y volviendo a calcular las sumas de cuadrados. Se debería llegar a que la suma de los cuadrados total y entre bloques se ha modificado, mientras que la suma de cuadrados residual y entre tratamientos no ha variado.

Si no hay ninguna diferencia entre las eficiencias o los días, entonces las tres sumas de cuadrados medios proporcionarían una estimación de  $\sigma_0^2$ , o sea, la varianza de la variación aleatoria debida al error experimental (véase la Sec. 3.9). La prueba  $F$  se utiliza para ver si las estimaciones de la varianza difieren significativamente. Comparando el cuadrado medio entre tratamientos con el cuadrado medio residual resulta:

$$F_{3,6} = 9.5556/1.6389 = 5.83$$

De la tabla A.2 podemos extraer el valor crítico de  $F$  que es 4.76 ( $P = 0.05$ ) y al ser superado por el valor anterior nos está indicando que hay una diferencia significativa entre tratamientos, es decir, entre la eficiencia de los diferentes agentes quelantes, al

nivel del 5%. Comparando el cuadrado medio entre bloques (es decir, entre días) con el residual resulta:

$$F_{2,6} = 7.75/1.6389 = 4.76$$

Como el valor crítico de  $F$  es 5.143 ( $P = 0.05$ ) y no es superado por el valor anterior, se concluye que no hay diferencias significativas entre días. No obstante, el cuadrado medio entre bloques es considerablemente mayor que el cuadrado medio residual, y si el experimento se hubiera realizado sin bloques, estos dos efectos se combinarían en la estimación del error experimental, y el experimento probablemente no habría sido capaz de detectar si tratamientos diferentes proporcionaban resultados significativamente diferentes. Si la diferencia entre días hubiera sido significativa, indicaría que otros factores tales como la temperatura, preparación de la disolución, etc., tendrían efectos. Se puede demostrar que el cuadrado medio entre bloques proporciona una estimación de  $\sigma_0^2 + c\sigma_e^2$ , donde  $\sigma_0^2$  es la varianza de la variación aleatoria de día a día. Ya que el cuadrado medio residual da una estimación de  $\sigma_0^2$ , se puede deducir una estimación de  $\sigma_e^2$ .

Este ejemplo ilustra claramente las ventajas de considerar el diseño de un experimento antes de realizarlo. Entre un experimento con bloques y uno sin bloques, con el mismo número de mediciones en cada uno, el primero es más sensible y proporciona más información.

El análisis expuesto supone que los efectos, si existen, son aditivos. Este aspecto se analiza posteriormente en la sección 7.7.

## 7.5 CUADRADOS LATINOS

Un factor incontrolado que no se consideró en el análisis de la sección anterior era el momento del día en que se realizaban las mediciones. Una variación sistemática durante el día debido al deterioro de la solución, o un cambio en una determinada dirección de la temperatura del laboratorio, podría producir una tendencia en los resultados. Es posible utilizar un diseño experimental que permita la separación de este factor si tiene el mismo número de bloques y tratamientos. Por ejemplo, del siguiente diseño de bloques aleatorizados

Día 1	A	C	B
Día 2	C	A	B
Día 3	C	B	A

en el que las mediciones que utilizan el tratamiento C se concentran al principio del día y las que usan el B lo hacen al final del día, se podría seguir el diseño siguiente:

Día 1	A	B	C
Día 2	C	A	B
Día 3	B	C	A

Este diseño, en el que cada tratamiento aparece una vez en cada fila y una vez cada columna, recibe el nombre de **cuadrado latino**; permite separar la variación entre tratamientos, entre días, entre el momento del día y el error experimental aleatorio.

Si el cuadrado es mayor que uno de  $3 \times 3$ , hay más de un diseño posible y se debería elegir al azar uno de estos diseños. La desventaja obvia de un diseño de cuadrado latino simple es que debe existir el mismo número de bloques y tratamientos. Es posible considerar diseños más complicados que eliminen esta restricción.

## 7.6 DISEÑOS ANIDADOS Y DE CLASIFICACIÓN CRUZADA

El diseño del experimento considerado en la sección 7.4 se denomina **clasificación cruzada** debido a que se toma una medición para cada combinación posible de los distintos niveles de los factores. En contraposición al diseño anterior se encuentra el diseño experimental que aparece en la figura 7.1, en el que una muestra de una solución estándar se envía a cada uno de los tres laboratorios y en cada laboratorio dos técnicos diferentes analizan la muestra. Se dice que este diseño es **anidado** o **jerárquico** y se puede ver que no se utilizan todas las combinaciones posibles de las condiciones experimentales, por ejemplo, cada técnico sólo realiza mediciones en su laboratorio. (El ANOVA simple se puede considerar también como un diseño anidado.) También es posible tomar en cuenta diseños mixtos (es decir, en parte de clasificación cruzada, en parte anidados).

## 7.7 INTERACCIÓN

La hipótesis básica en el análisis realizado en la sección 7.4 era que los efectos de los dos factores fuesen **aditivos**. El significado de este término se explica mejor mediante un ejemplo numérico. La situación más simple es aquella en la que cada factor tiene dos niveles: por ejemplo, el factor A tiene los niveles  $a_1$  y  $a_2$ , y el factor B los niveles  $b_1$  y  $b_2$ . Supongamos que las combinaciones  $a_1b_1$ ,  $a_1b_2$  y  $a_2b_1$  tienen como resultados 10, 15 y 12 como se muestra en la tabla 7.3.

Si éstos fuesen verdaderos valores, es decir, si no existiese ninguna variación aleatoria, entonces el efecto de cambiar el factor B del nivel  $b_1$  al nivel  $b_2$ , manteniendo A en  $a_1$ , sería 5. Si los efectos fuesen aditivos, entonces el efecto de cambiar B del nivel  $b_1$  al nivel  $b_2$

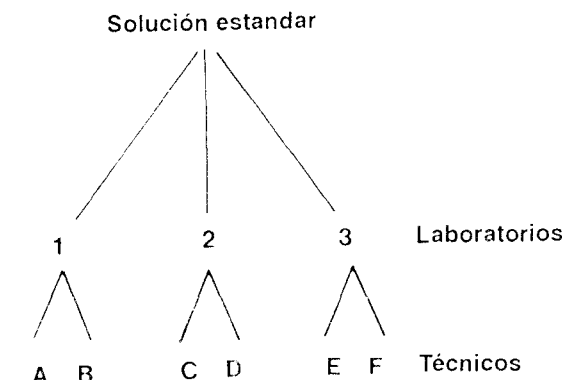


Figura 7.1 Diseño anidado o jerárquico.

Tabla 7.3

	Nivel de A	
	$a_1$	$a_2$
Nivel de B	$b_1$	10
	$b_2$	15

con A en  $a_2$  también debería ser 5, de manera que el valor final en la tabla debería ser 17. Observe que entonces el efecto de cambiar A del nivel  $a_1$  al nivel  $a_2$  es 2, independientemente del nivel de B. En general, si los resultados son como se muestran en la tabla 7.4 y son ajenos al error aleatorio, entonces  $y_2 - y_1$  es igual a  $y_4 - y_3$  si los efectos son

Tabla 7.4

	Nivel de A	
	$a_1$	$a_2$
Nivel de B	$b_1$	$y_1$
	$b_2$	$y_2$

aditivos. Si no es así, se dice que hay **interacción** entre A y B, ya que en una combinación concreta entre A y B da un valor más grande que el esperado. La figura 7.2 aclara gráficamente el efecto de la interacción. En las dos gráficas cada resultado se representa respecto al nivel de B, y se unen los puntos que tienen el mismo nivel de A. Si los efectos fuesen aditivos, entonces las rectas serían paralelas como en el diagrama de la parte

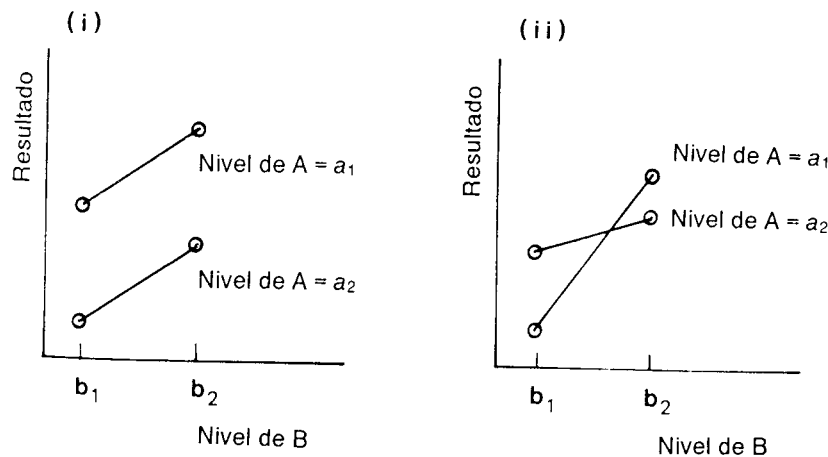


Figura 7.2 (i) Efectos que son aditivos; (ii) efectos que interaccionan.

izquierda; si los efectos interaccionasen, las rectas no serían paralelas, como en el diagrama de la derecha. Si los factores tienen varios niveles, una gráfica de este tipo es útil en la interpretación de cualquier interacción.

Desafortunadamente la situación experimental se confunde por la presencia de errores aleatorios. El lector puede comprobar que para la tabla 7.3, un ANOVA de dos factores da cero para la suma de cuadrados residual si el valor faltante se considera como 17, pero si uno de los valores se modifica, esto ya no es así. Con este diseño del experimento no podemos valorar hasta que punto una suma de cuadrados residual que no es cero es causada por el error aleatorio, en lugar de serlo por la interacción entre los dos efectos. Para estimar el error aleatorio se debe repetir las mediciones en cada celda de la tabla. Es importante la forma exacta de esta repetición. Supongamos, por ejemplo, que una determinación consiste en disolver un peso dado de una especie en un volumen especificado de agua y posteriormente en titularlo frente a una disolución estándar. Las mediciones repetidas deben incluir tanto la etapa de pesar como la de titular: si sólo se repitiese la etapa de titular, el error aleatorio producido al pesar no estaría incluido en la estimación del error experimental. Además, si el mismo lote de disolución estándar y el mismo conjunto de material de vidrio no se utilizan para *todas* las determinaciones en el experimento, entonces *no* se debe utilizar el mismo lote de disolución estándar y el mismo conjunto de material de vidrio para las mediciones repetidas. Las mediciones que se repiten de manera que están sujetas a todas las fuentes de error aleatorio en el experimento se llaman **réplicas**. En el ejemplo siguiente se aclara el método por el cual se puede separar el error aleatorio y la suma de cuadrados de la interacción.

*Ejemplo.* En un experimento para investigar la validez de una solución como estándar de absorbancia, se midió el valor de la absorptividad molar,  $\epsilon$ , de soluciones de tres concentraciones diferentes a cuatro longitudes de onda diferentes. Se realizaron dos mediciones repetidas por cada combinación de concentración y longitud de onda; y se aleatorizó el orden en que se hicieron las mediciones. Los resultados se muestran en la tabla 7.5: para simplificar los cálculos se han multiplicado por 100 los valores medidos.

Tabla 7.5

Concentración (g/l) \ $\lambda$ (nm)	240	270	300	350
0.02	94, 96	106, 108	48, 51	78, 81
0.06	93, 93	106, 105	47, 48	78, 78
0.10	93, 94	106, 107	49, 50	78, 79

El primer paso en los cálculos es encontrar los totales de celda. Esto se hace en la tabla 7.6, que también incluye otras cantidades necesarias en los cálculos. Como antes,  $T_i$  denota el total de la fila  $i$ -ésima, el  $T_j$  el total de la columna  $j$ -ésima y  $T$  el gran total.

Se calculan las sumas de cuadrados entre filas, entre columnas y del total. Cada cálculo requiere el término  $T^2/nrc$  (donde  $n$  es el número de mediciones repetidas en

Tabla 7.6

Concentración (g/l) \ λ(nm)	240	270	300	350	$T_{i.}$	$T_i^2$
0.02	190	214	99	159	662	438244
0.06	186	211	95	156	648	419904
0.10	187	213	99	157	656	430336
$T_{.j}$	563	638	293	472	$T = 1966$	$1288484 = \sum_i T_i^2$
$T_j^2$	316969	407044	85849	222784		
	$\sum_j T_j^2 = 10322646$					

cada celda, en este caso 2,  $r$  = número de filas y  $c$  = número de columnas). A este término se le denomina a veces **término de corrección**,  $C$ . Tenemos:

$$C = T^2/nrc = 1966^2/(2 \cdot 3 \cdot 4) = 161048.16$$

Se calculan ahora las sumas de cuadrados:

$$\begin{aligned} \text{Suma de cuadrados entre filas} &= \sum_i T_i^2 / nc - C \\ &= 1288484 / (2 \times 4) - 161048.16 \\ &= 12.34 \end{aligned}$$

con  $r - 1 = 2$  grados de libertad

$$\begin{aligned} \text{Suma de cuadrados entre columnas} &= \sum_j T_j^2 / nc - C \\ &= 1032646 / (2 \cdot 3) - 161048.16 \\ &= 11059.506 \end{aligned}$$

con  $c - 1 = 3$  grados de libertad.

$$\text{Suma de cuadrados total} = \sum x_{ijk}^2 - C$$

donde  $x_{ijk}$  es la  $k$ -ésima réplica en la  $i$ -ésima fila y  $j$ -ésima columna, es decir,  $\sum x_{ijk}^2$  es la suma de los cuadrados de las mediciones individuales que aparecen en la tabla 7.5.

$$\begin{aligned} \text{Suma de cuadrados total} &= 172138 - 161048.16 \\ &= 11089.84 \end{aligned}$$

con  $nrc - 1 = 23$  grados de libertad.

La variación debida al error aleatorio (normalmente llamada **variación residual**) se estima a partir de la variación dentro de las celdas, es decir, la variación entre réplicas. La suma de cuadrados residual =  $\sum x_{ijk}^2 - \sum T_{ij}^2 / n$  donde  $T_{ij}$  es el total para la celda

de la  $i$ -ésima fila y  $j$ -ésima columna, es decir, la suma de las mediciones replicadas en la  $i$ -ésima fila y la  $j$ -ésima columna.

$$\begin{aligned} \text{Suma de cuadrados residual} &= \sum x_{ijk}^2 - \sum T_{ij}^2 / n \\ &= 172138 - 344244 / 2 \\ &= 16 \end{aligned}$$

con  $(n - 1)rc = 12$  grados de libertad.

La suma de cuadrados de la interacción y sus grados de libertad se pueden obtener ahora por diferencia. A continuación se resumen los resultados de estos cálculos.

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio
Entre filas	12.34	2	6.17
Entre columnas	11059.506	3	3686.502
Interacción	1.994	6	0.3323
Residual	16	12	1.3333
Total	11089.84	23	

Para probar si cada una de las fuentes de variación es significativa se compara el cuadrado medio de dicha fuente con el cuadrado medio residual.

(i) **Interacción**. Ésta es obviamente no significativa ya que el cuadrado medio de la interacción es menor que el cuadrado medio residual.

(ii) **Entre columnas** (es decir, entre longitudes de onda). Ésta es significativa ya que tenemos:

$$F_{3,12} = 3686.502 / 1.3333 = 2765$$

El valor crítico de  $F_{3,12}$  es 3.49 ( $P = 0.05$ ). En este caso se esperaba un resultado significativo ya que la absorbancia depende de la longitud de onda.

(iii) **Entre filas** (es decir, entre concentraciones). Tenemos:

$$F_{2,12} = 6.17 / 1.3333 = 4.63$$

El valor crítico de  $F_{2,12}$  es 3.885 ( $P = 0.05$ ), indicando que la variación entre filas es demasiado grande para que se tenga en cuenta como variación aleatoria: la solución no es adecuada como un estándar de absorbancia. La figura 7.3 muestra la representación de la absorptividad molar frente a la longitud de onda, y en ella los valores con la misma concentración están unidos por líneas rectas. Esto aclara los resultados del análisis anterior de la siguiente forma:

- (i) las líneas son paralelas, lo que indica ausencia de interacción;
- (ii) las líneas no son muy horizontales, lo que indica que la absorptividad molar varía con la concentración;
- (iii) las líneas se encuentran a diferentes alturas en la gráfica, lo que indica que la absorptividad molar depende de la longitud de onda.

Las fórmulas en los cálculos anteriores se resumen en la tabla 7.7.

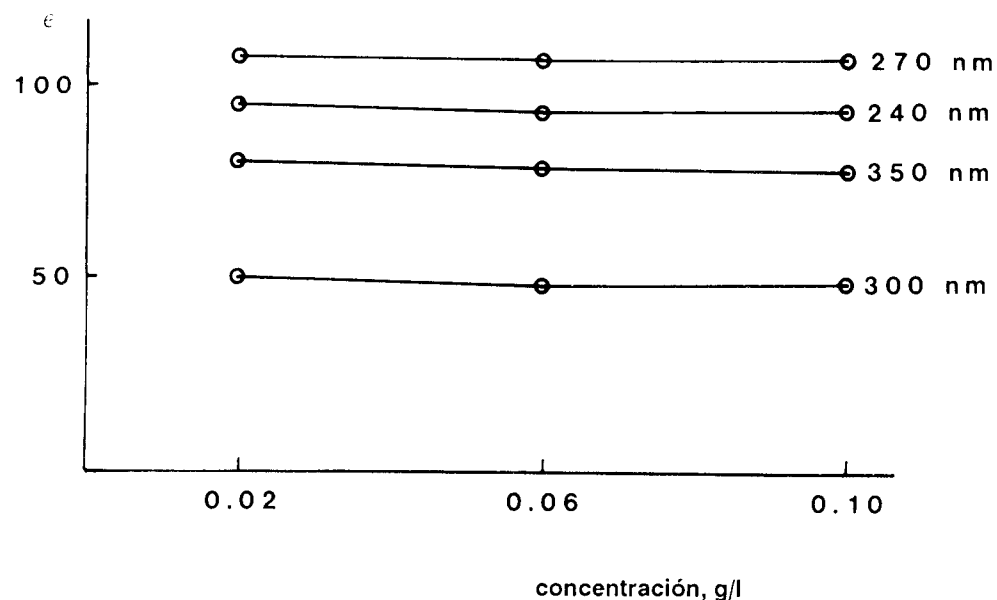


Figura 7.3 Relaciones en el ANOVA de dos factores del ejemplo de la sección 7.7.

En este experimento los dos factores, es decir, la longitud de onda y la concentración de la solución, son factores controlados. En química analítica una aplicación importante de las técnicas ANOVA es investigar dos o más factores controlados y sus interacciones en experimentos de optimización. Esto se trata en la sección 7.9.

Como se comentó en la sección 4.8, otra aplicación importante de las técnicas ANOVA se encuentran en las investigaciones de colaboración entre laboratorios, en cuanto a precisión y exactitud. Una versión más elaborada de los experimentos descritos en la sección 4.8 supondría enviar varios tipos diferentes de muestras a una serie de laboratorios y que cada uno realizase una serie de análisis replicados sobre cada muestra. Un análisis matemático de los resultados proporcionaría las siguientes sumas de cuadrados: entre laboratorios, entre muestras, interacción muestra-laboratorio, y residual. El objetivo de dicho experimento sería probar primero si hay interacción entre laboratorio y muestra, es

Tabla 7.7 Fórmulas para ANOVA de dos factores con interacción

Fuente de variación	Suma de cuadrados	Grados de libertad
Entre filas	$\sum_i T_i^2 / nc - C$	$r - 1$
Entre columnas	$\sum_j T_j^2 / nc - C$	$c - 1$
Interacción	por diferencia	por diferencia
Residuos	$\sum x_{ijk}^2 - \sum T_{ij}^2 / n$	$rc(n - 1)$
Total	$\sum x_{ijk}^2 - C$	$rcn - 1$

decir, si algunos laboratorios daban resultados inesperadamente altos o bajos para algunas muestras. Esto se hace comparando la suma de cuadrados residual y de interacción. Si *no* hay ninguna interacción, entonces se podría probar si los laboratorios obtuvieron resultados significativamente diferentes, es decir, si existe diferencia sistemática entre los laboratorios. Si existiese, entonces se podría estimar la varianza entre ellos. Sin embargo, si *hay* alguna interacción significativa, la prueba sobre diferencias significativas entre laboratorios tiene poca relevancia.

Para que un ANOVA de dos factores sea válido se deben satisfacer las siguientes condiciones (véase la Sec. 3.9):

- (i) el error aleatorio es el mismo para todas las combinaciones de niveles de los factores;
- (ii) los errores aleatorios se distribuyen aproximadamente normales.

## 7.8 EL DISEÑO FACTORIAL FRENTE AL DE UN FACTOR CADA VEZ

Un experimento como el del ejemplo anterior, en el que la variable de respuesta (es decir, la absorbividad molar) se mide para todas las combinaciones posibles de los niveles elegidos de los factores, recibe el nombre de **diseño factorial completo**. El lector ha podido observar que este tipo de diseño de experimento es la antítesis de la aproximación clásica en donde la respuesta se investiga para cada factor, mientras que el resto de los factores se mantiene a un nivel constante. Hay dos razones de peso para utilizar un diseño factorial en lugar de uno clásico, en experimentos que prueben si la respuesta depende del nivel del factor:

(i) el experimento factorial detecta y estima cualquier interacción que el experimento de un factor cada vez no puede hacer;

(ii) si los efectos de los factores son aditivos, entonces el diseño factorial necesita menos mediciones que la aproximación clásica para proporcionar la misma precisión. Esto se puede ver volviendo nuevamente al experimento de absorbividad molar. Allí, las 24 mediciones se utilizaban para estimar el efecto de variar la longitud de onda y las *mismas* 24 para estimar el efecto de variar la concentración.

En un experimento de un factor cada vez, en primer lugar se habría fijado la concentración y, para obtener la misma precisión para el efecto de variar la longitud de onda, habrían sido necesarias 6 mediciones para cada longitud de onda, es decir, 24 en total. Después se habría fijado la longitud de onda y se realizarían otras 24 mediciones para concentraciones diferentes, formando en total 48. En general, para  $k$  factores, una aproximación clásica supone  $k$  veces tantas mediciones como un factorial con la misma precisión.

## 7.9 DISEÑO FACTORIAL Y OPTIMIZACIÓN

Puesto que con frecuencia las técnicas analíticas están interesadas en la detección de trazas de analito, suele ser importante que los niveles de los factores de los que depende la respuesta se elijan con vistas a maximizar la respuesta. El proceso de búsqueda de estos niveles de factor óptimos se conoce como **optimización**.

En el primer paso se determinan los factores y las interacciones que afectan de manera importante a la respuesta. Esto se puede hacer utilizando un experimento factorial en el



que cada factor tenga dos niveles, conocidos habitualmente por “alto” y “bajo”. En el caso de una variable cuantitativa los términos “alto” y “bajo” tiene el significado usual. La elección exacta de los niveles se determina principalmente por la experiencia y el conocimiento del experimentador y las restricciones físicas del sistema, por ejemplo, si el agua se utiliza como disolvente, sólo se puede trabajar con temperaturas en el intervalo 0-100°C. Algunos problemas relacionados con la elección de los niveles se analizan más adelante. Para una variable cualitativa, “alto” y “bajo” se refiere a un par de condiciones diferentes, por ejemplo, la presencia o ausencia de un catalizador, el uso de agitación mecánica o magnética, muestras de forma granular o el polvo, etc. Como ya se han considerado experimentos con dos factores con cierto detalle, nos centraremos en uno con tres factores A, B y C. Esto significa que hay  $2 \times 2 \times 2 = 8$  combinaciones posibles de niveles de factores, como se muestran en la tabla siguiente. Un signo más indica que el factor está en el nivel alto y uno menos que está en el nivel bajo. La primera columna muestra una notación utilizada a menudo para describir las combinaciones, donde la presencia de la correspondiente letra en minúscula indica que el factor está en el nivel alto, y su ausencia, que el factor está en el nivel bajo. El número 1 se utiliza para indicar que todos los factores están en el nivel bajo.

Combinación	A	B	C	Respuesta
1	-	-	-	$y_1$
a	+	-	-	$y_2$
b	-	+	-	$y_3$
c	-	-	+	$y_4$
bc	-	+	+	$y_5$
ac	+	-	+	$y_6$
ab	+	+	-	$y_7$
abc	+	+	+	$y_8$

El método por el cual se estiman los efectos de los factores y sus interacciones se ilustra mediante el ejemplo siguiente.

*Ejemplo.* En un experimento de cromatografía líquida de alta resolución, se investigó la dependencia del parámetro de retención,  $k'$ , sobre tres factores. Los factores fueron pH (factor P), la concentración de un contraión (factor T) y la concentración de un disolvente orgánico en la fase móvil (factor C). Se utilizaron dos niveles para cada factor y se realizaron dos mediciones, una réplica de la otra, para cada combinación. Las mediciones se aleatorizaron. La tabla que se expone a continuación da el valor medio para cada par de réplicas.

Combinación de niveles de factor	$k'$
1	4.7
p	9.9
t	7.0
c	2.7
pt	15.0
pc	5.3
tc	3.2
ptc	6.0

### (a) Efecto de los factores individuales.

El efecto de cambiar los niveles de P se puede hallar de la diferencia media en respuesta cuando P cambia el nivel alto al bajo y los niveles de C y T se mantienen fijos. Hay cuatro pares de respuestas que proporcionan una estimación del efecto del nivel P, como se muestra en la tabla siguiente.

Nivel de C	Nivel de T	Nivel de P		Diferencia
		+	-	
-	-	9.9	4.7	5.2
+	-	5.3	2.7	2.6
-	+	15.0	7.0	8.0
+	+	6.0	3.2	2.8
Total = 18.6				

Efecto promedio de alterar el nivel de P =  $18.6/4 = 4.65$

De manera similar se puede obtener el efecto promedio de alterar los niveles de T y C. Se deja como ejercicio para el lector demostrar que ellos son:

efecto promedio de alterar el nivel de C = - 4.85

efecto promedio de alterar el nivel de T = 2.15

### (b) Interacción entre dos factores.

Consideremos ahora los dos factores P y T. Si *no* hubiera interacción entre ellos, entonces el cambio de respuesta entre los dos niveles de P debería ser independiente del nivel T. Las primeras dos cifras en la última columna de la tabla anterior dan el cambio en respuesta cuando P cambia del nivel alto al bajo con T en el nivel bajo. Su promedio es  $(5.2 + 2.6)/2 = 3.9$ . Las dos últimas cifras en la misma columna proporcionan el efecto de cambiar P cuando T está en el nivel alto. Su promedio es  $(8.0 + 2.8)/2 = 5.4$ . Si no hubiera interacción ni error aleatorio (véase la Sec 7.7) estas estimaciones del efecto del cambio del nivel P deberían ser iguales. El convenio seguido es tomar la mitad de su diferencia como una medida de la interacción:

efecto de interacción PT =  $(5.4 - 3.9)/2 = 0.75$ .

Es importante darse cuenta de que esta cantidad estima el grado en que los efectos de P y T no son aditivos. Igualmente, se podría haber calculado considerando qué tanto el cambio en respuesta para los dos niveles de T es independiente del nivel de P.

Las otras interacciones se calculan de manera similar. De nuevo el lector puede verificar que:

efecto de interacción CP = - 1.95

efecto de interacción CT = - 1.55

### (c) Interacción entre tres factores.

La interacción PT calculada anteriormente se puede dividir en dos partes con relación al nivel de C. Con C en el nivel bajo, la estimación de la interacción sería  $(8.0 - 5.2)/2$

= 1.4 y con C en el nivel alto sería  $(2.8 - 2.6)/2 = 0.1$ . Si no hubiera interacción entre los tres factores ni error aleatorio, estas estimaciones de la interacción PT deberían ser iguales. La interacción de los tres factores se estima por la mitad de su diferencia [ $= (0.1 - 1.4)/2 = -0.65$ ]; y mide hasta que punto el efecto de la interacción PT y el efecto de C no son aditivos: igualmente se podría calcular considerando la diferencia entre las estimaciones de la interacción PC para los niveles bajo y alto de T, o la diferencia entre las estimaciones de la interacción TC para los niveles bajo y alto de P.

Estos resultados se resumen en la tabla siguiente.

	Efecto
Un único factor (efecto principal)	
P	4.65
T	2.15
C	-4.85
Interacción de dos factores	
TP	0.75
CT	-1.55
CP	-1.95
Interacción de tres factores	
PTC	-0.65

Los cálculos se han presentado con cierto detalle para mostrar con más claridad los principios. Un algoritmo debido a Yates (véanse la referencias) simplifica el cálculo.

Para probar qué efectos, si existen, son significativos, se puede utilizar el ANOVA (suponiendo que existe homogeneidad de varianza). Se puede demostrar que en un experimento de *dos* niveles, como éste, las sumas de cuadrados necesarias se pueden calcular de los efectos estimados utilizando

$$\text{Suma de cuadrados} = N \times (\text{efecto estimado})^2/4$$

donde  $N$  es el número total de mediciones incluyendo las réplicas. En este caso  $N$  es 16, ya que se hicieron dos mediciones, una réplica de la otra, por cada combinación de los niveles del factor. Las sumas de cuadrados calculadas se proporcionan a continuación

Factor (es)	Suma de cuadrados
P	86.49
T	18.49
C	94.09
PT	2.25
TC	9.61
PC	15.21
PTC	1.69

Se puede demostrar que cada suma de cuadrados tiene un grado de libertad, y puesto que el cuadrado medio está dado por:

$$\text{cuadrado medio} = \text{suma de cuadrados} / \text{número de grados de libertad}$$

cada cuadrado medio es simplemente la correspondiente suma de cuadrados. Para contrastar si un efecto es significativo, se compara el cuadrado medio con el cuadrado

medio del error (residual). Éste se calcula de las mediciones individuales realizadas por el método descrito en el ejemplo de absortividad molar en la sección 7.7. En el experimento que estamos analizando, se ha obtenido 0.012 para el cuadrado medio residual con 8 grados de libertad. El contraste de significación, empezando con la interacción de orden más grande, proporciona para la interacción PTC:

$$F_{1,8} = 1.69/0.012 = 141$$

que es obviamente significativa. Si existe interacción entre los tres factores, no tiene caso probar que los factores tomados por pares o individualmente sean significativos, ya que en cualquier proceso de optimización se tendrán que considerar todos los factores. Sólo se debería probar la significación de un único factor si no hubiera interacción con otros factores.

Un problema que aparece en un experimento factorial completo como éste es que el número de experimentos requerido crece rápidamente con el número de factores: para  $k$  factores a 2 niveles con 2 réplicas para cada combinación de valores, son necesarios  $2^{k+1}$  experimentos, por ejemplo, para 5 factores, 64 experimentos. Cuando hay más de tres factores es posible una simplificación suponiendo que las interacciones de orden 3 y superiores sean despreciables. Las sumas de cuadrados correspondientes a estas interacciones se pueden combinar para proporcionar una estimación de la suma de cuadrados residual y que no sean necesarias más réplicas. La lógica para esta aproximación es que los efectos de orden más grande son normalmente mucho más pequeños que los efectos principales y los efectos de interacción entre dos factores. Si las interacciones de orden superior *se pueden* suponer insignificantes, una fracción adecuada de todas las posibles combinaciones de los niveles del factor es suficiente para proporcionar una estimación de los efectos principal y de interacción de dos factores. Como se mencionó en la sección 4.7, dicho diseño experimental se llama **diseño factorial fraccional o incompleto**.

Al utilizar un diseño factorial para determinar qué factores tienen un efecto significativo sobre la respuesta se observa que, para factores que son variables continuas, el efecto depende de los niveles alto y bajo utilizados; si éstos están excesivamente cerca uno del otro, se puede obtener que el efecto del factor correspondiente sea no significativo, a pesar de que sobre el intervalo posible completo de los niveles del factor, el efecto de ese factor *no* sea despreciable. Por otra parte, si los niveles se alejan bastante pueden caer a ambos lados de un máximo, y aun así dar lugar a una diferencia en la respuesta que no sea significativa.

## 7.10 INCAPACIDAD DEL MÉTODO DE UN FACTOR CADA VEZ PARA LOGRAR EL ÓPTIMO

En la sección 7.8 se demostró la superioridad de un diseño factorial sobre un experimento de un factor cada vez en el momento de estimar si un factor ejercía un efecto significativo. El método de un factor cada vez es también poco adecuado para lograr condiciones óptimas si la respuesta depende de dos o más factores que interaccionan (véase la Sec. 1.6). Esto se puede comprobar gráficamente cuando la respuesta depende de dos factores que son variables continuas. En este caso, la relación entre la respuesta y los niveles de los dos

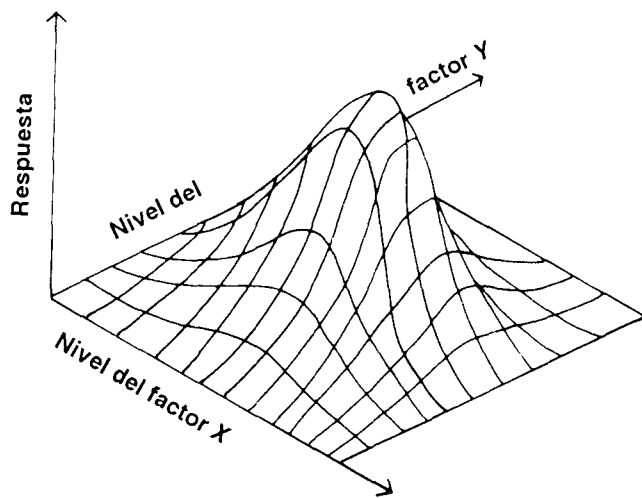


Figura 7.4 Una superficie de respuesta para dos factores.

factores se puede representar por una superficie tridimensional como se muestra en la figura 7.4. Esta superficie se llama **superficie de respuesta**. Se comprende mejor su utilidad considerando un diagrama de curvas de nivel (véase la Fig. 7.5), en el que la respuesta sobre cada curva de nivel es constante. La forma de las curvas de nivel es, desde luego, desconocida por el investigador, quien desea determinar los niveles óptimos,  $x_0$  y  $y_0$  para los factores  $X$  y  $Y$ , respectivamente. Adoptando el enfoque de un factor cada vez con el nivel inicial de  $X$  fijo en  $x_1$ , por ejemplo, y variando el nivel de  $Y$ , resultaría una respuesta máxima en el punto A, cuando el nivel de  $Y$  fuese  $y_1$ . A continuación, manteniendo el nivel de  $Y$  en  $y_1$ , y variando el nivel de  $X$  resultaría un máximo en B. Obviamente, éste no es el máximo real y la posición alcanzada depende del nivel inicial elegido para  $x_1$ . El proceso se podría repetir variando los niveles de  $X$  y  $Y$  alternativamente, pero éste es un método no eficiente de aproximación a  $(x_0, y_0)$ . Además, en algunos casos no conduciría al máximo real, como se pone de manifiesto en la figura 7.6 donde, aunque C no es el máximo real, el valor de respuesta disminuye hacia cada lado de las direcciones  $X$  y  $Y$ . Con el método de un factor cada vez, al llegar a este punto se concluiría, por tanto, que representaba la respuesta máxima. Razonamientos similares se aplican en la optimización de más de dos factores, aunque no se pueda visualizar fácilmente la superficie de respuesta en estos casos. Por esta razón, el análisis de los métodos de optimización se centrará en gran parte en el caso de dos factores, aunque los métodos se puedan generalizar a cualquier número de factores.

## 7.11 MÉTODO DE LA MÁXIMA PENDIENTE O DEL GRADIENTE

El proceso de optimización se puede visualizar como un hombre en una colina con niebla espesa y que pretende llegar a la cima. En estas circunstancias una aproximación obvia es caminar en la dirección en que la colina tenga mayor pendiente. Ésta es la base del **método del gradiente**. La figura 7.7 muestra dos mapas posibles de curvas de nivel. Un simple razonamiento demuestra que la dirección de máxima pendiente en cualquier punto se

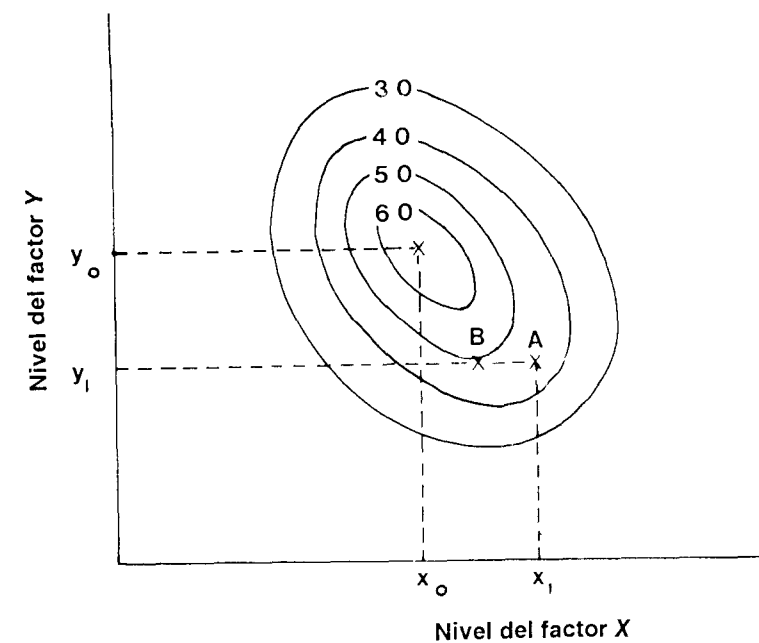


Figura 7.5 Diagrama de las curvas de nivel para una superficie de respuesta con dos factores.

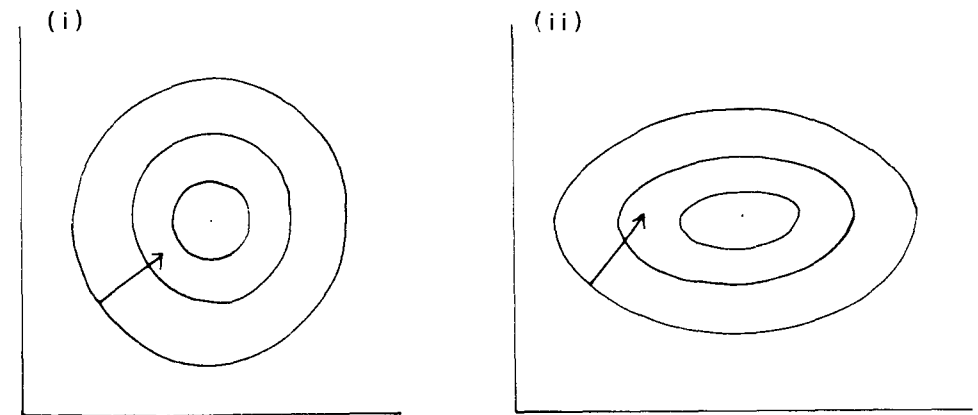
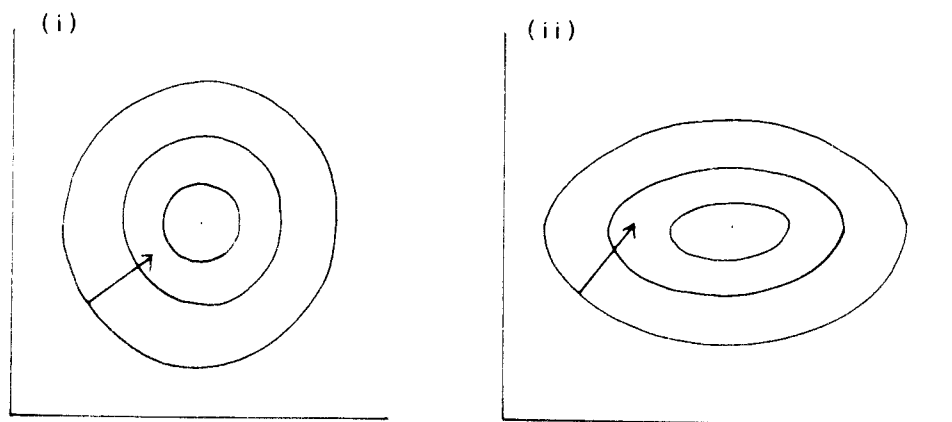


Figura 7.6 Diagrama de curvas de nivel: un caso en el que el método de un factor cada vez no sirve para localizar el máximo.



**Figura 7.7** Diagramas de las curvas de nivel: la flecha en cada diagrama indica el camino de la máxima pendiente. En (i) se dirige aproximadamente al máximo pero en (ii) no.

encuentra en ángulo recto a las curvas de nivel en dicho punto, como se indica con las flechas. Cuando las curvas de nivel son círculos, las flechas apuntarán a la cima, pero cuando sean elipses puede no cumplirse lo anterior. La forma de las curvas de nivel depende de las escalas elegidas para los ejes: se obtienen los mejores resultados con este método si se toma una escala de los ejes de manera que un cambio de una unidad en cualquier dirección dé un cambio aproximadamente igual en la respuesta. El primer paso es realizar un experimento factorial con cada factor a dos niveles. Los niveles se eligen de tal manera que el diseño forme un cuadrado como el que se muestra en la figura 7.8. Suponga, por ejemplo, que el experimento es una reacción (catalizada por enzimas) en la que la tasa de reacción, en este caso es la respuesta, se quiere maximizar respecto del pH (factor X) y la temperatura (factor Y). La tabla siguiente proporciona los resultados (tasa de reacción medida en unidades arbitrarias) del experimento factorial inicial.

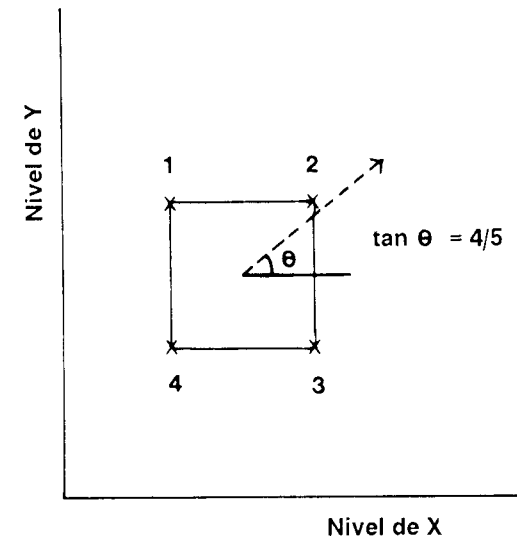
		pH (X)	
		6.8	7.0
Temperatura (°C)	20	30	35
	25	34	39

Los efectos de los dos factores se pueden separar según se describió en la sección 7.9. Escribiendo de nuevo la tabla anterior, con la notación de esa sección, resulta:

Combinación de niveles	Tasa de reacción
1	30
x	35
y	34
xy	39

$$\text{Efecto promedio del cambio en el nivel de X} = [(35 - 30) + (39 - 34)]/2 = 5$$

$$\text{Efecto promedio del cambio en el nivel de Y} = [(34 - 30) + (39 - 35)]/2 = 4$$



**Figura 7.8** Un diseño factorial  $2 \times 2$  para determinar la máxima pendiente, que está indicada por la línea de trazo discontinuo.

Los efectos de X y de Y indican que en la figura 7.8 se buscaría la máxima respuesta a la derecha y por encima de la región original. Ya que la variación en la dirección de X es mayor que en la dirección de Y, el incremento a dar en la dirección de la X tendría que ser mayor. Para ser más exactos, los incrementos en las direcciones de X y Y tendrían que estar en razón de 5 a 4 respectivamente, es decir, en la dirección indicada por la línea de trazo discontinuo de la figura 7.9.

El paso siguiente en la optimización es realizar otros experimentos en la dirección indicada por la línea discontinua en la figura 7.9, digamos que en los puntos cuya numeración es 5, 6 y 7. Esto indicaría que el punto 6 es una posición aproximada para el máximo en esta dirección. Entonces se lleva a cabo otro experimento factorial en la región del punto 6 para determinar la nueva dirección de la máxima pendiente.

Este método proporciona una forma satisfactoria de progresar en dirección al máximo, con tal que sobre la región del diseño factorial, las curvas de nivel sean aproximadamente rectas. Esto es equivalente a que la superficie de respuesta sea plana, lo que se puede describir matemáticamente por una combinación lineal de términos en x y y. Al estar más cerca de la cumbre son también necesarios términos en xy,  $x^2$  y  $y^2$ . El término xy representa la interacción entre X y Y y se puede estimar utilizando la replicación como se describió en la sección 7.9. Los términos cuadrados, que representan la curvatura de la superficie, se pueden estimar comparando la respuesta en el centro del diseño factorial con el promedio de las respuestas en las esquinas. Cuando los efectos de curvatura e interacción se hacen más apreciables en comparación con el error experimental (que se estima mediante replicación) se emplea un diseño factorial más elaborado que permite determinar la forma de la superficie curva y, por lo tanto, la posición aproximada del máximo.

Es evidente que el diseño factorial y el método del gradiente serán muy complicados cuando existan varios factores implicados. La siguiente sección describe un método de optimización que es conceptualmente mucho más simple.

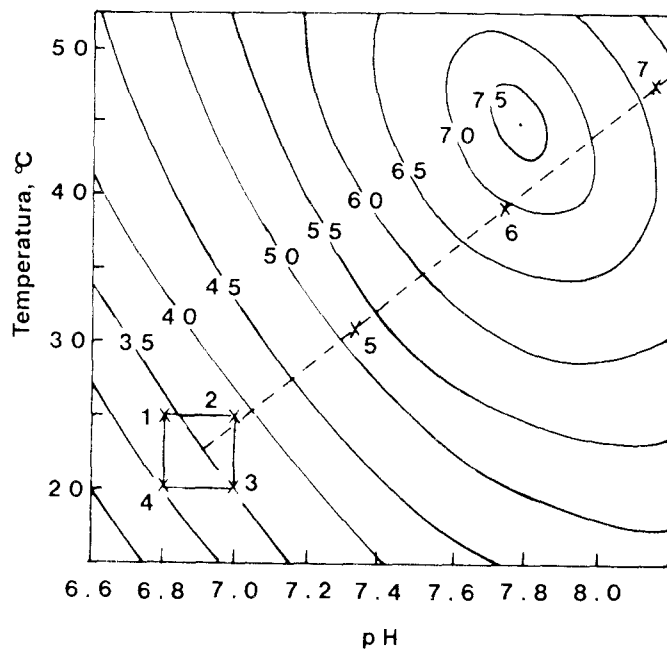


Figura 7.9 Diagrama de curvas de nivel: la dirección inicial de la máxima pendiente es la correspondiente a la línea de trazo discontinuo. En los puntos 5, 6 y 7 se realizan nuevos experimentos.

## 7.12 OPTIMIZACIÓN POR EL MÉTODO SIMPLEX

La optimización simplex se aplica cuando todos los factores son variables continuas. (Se advierte al lector que la elección del nombre es poco afortunada ya que el método es diferente del método simplex que se utiliza en programación lineal.) Un **simplex** es una figura geométrica que tiene  $n + 1$  vértices cuando se trata de optimizar una respuesta respecto a  $n$  factores. Por ejemplo, para dos factores, el simplex será un triángulo. El método de optimización se ilustra mediante la figura 7.10. El simplex inicial está definido por los puntos con numeración 1, 2 y 3. En el primer experimento se mide la respuesta en cada una de las combinaciones de los niveles del factor dadas por los vértices del simplex. En este caso, la peor respuesta se encontraría en el punto 3 y sería lógico concluir que se obtendría una respuesta mejor en el punto simétrico del 3 respecto a la línea que une 1 y 2, es decir, en 4. Con los puntos 1, 2 y 4 se forma un nuevo simplex y se mide la respuesta para la combinación de los niveles del factor dada por 4. Si ahora se comparan las respuestas alcanzadas en los puntos 1, 2 y 4 se observará que 1 da la peor respuesta. El procedimiento anterior, basado en la reflexión sobre una recta, se repite de nuevo, con lo cual el simplex resulta definido por los puntos 2, 4 y 5. La figura 7.10 muestra la continuación de este proceso. Se puede ver que no es posible progresar más allá del paso dado, ya que los puntos 6 y 8 dan peor respuesta que 5 y 7.

Para mejorar la ejecución del método simplex se han propuesto (y se siguen proponiendo) varias modificaciones. El avance hacia el máximo se puede acelerar utilizando un simplex que puede variar en tamaño según como se compare la respuesta en el nuevo vértice en un simplex con los otros vértices. Inicialmente, el simplex es grande para avanzar con

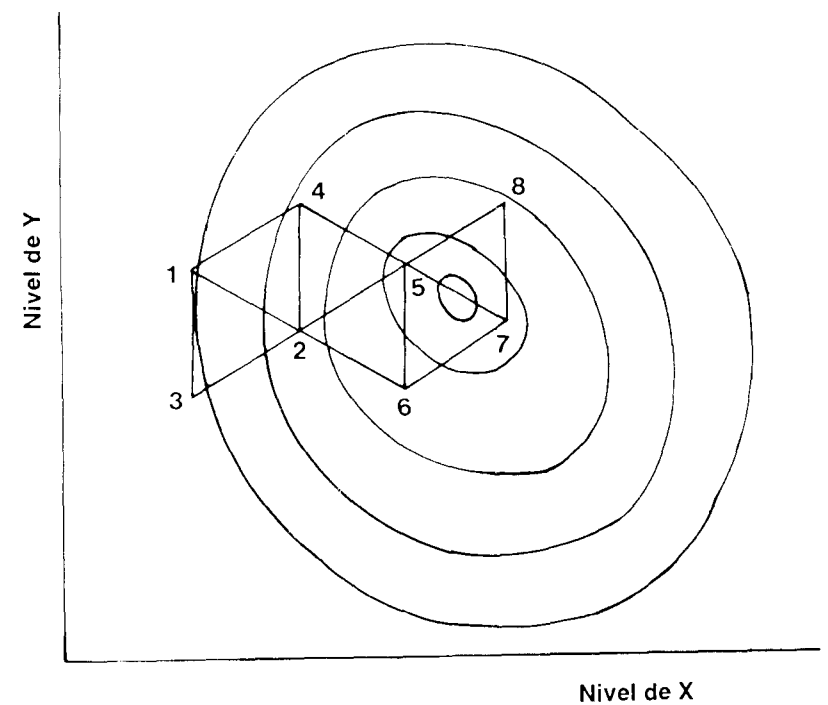


Figura 7.10 Optimización por el método simplex.

rapidez hacia el máximo; cerca del máximo se contrae para permitir que el máximo se pueda hallar con exactitud y evitar la situación mostrada en la figura 7.10.

La posición del nuevo vértice de un simplex se puede encontrar por cálculo en lugar de dibujar —esto es esencial cuando hay más de dos factores. Los cálculos se establecen con más facilidad como se muestra en la tabla 7.8. En este ejemplo hay 5 factores y, por tanto, el simplex tiene 6 vértices. La respuesta para el vértice 4 es la más pequeña y en consecuencia se tiene que sustituir este vértice. Las coordenadas del centro de gravedad de los vértices que se *conservan* se calculan de la siguiente forma: (i) sumando las coordenadas de los vértices que se conservan y (ii) dividiendo por el número de factores,  $n$ . El desplazamiento del nuevo punto desde el centro del gravedad está dado por (vi) = (ii) - (iii), y las coordenadas del nuevo vértice, vértice 7, por (v) = (ii) + (iv). Si se varía el simplex en tamaño, entonces los valores en la fila (iv) se multiplican por un factor de escala adecuado.

Una cuestión obvia al utilizar el método del simplex es la elección del simplex inicial. Si éste se considera como una figura *regular* en  $n$  dimensiones, entonces las posiciones de los vértices para generar tal figura dependerán de las escalas usadas en los ejes. Como en el método de la máxima pendiente o del gradiente, estas escalas se deberían elegir de manera que el cambio de unidad en cada factor dé aproximadamente el mismo cambio en la respuesta. Si hay insuficiente información para lograr esto, la diferencia entre los valores mayor y menor factibles de cada factor se puede representar por la misma distancia. La elección del tamaño del simplex inicial se analiza en un artículo de Yarbrow y

**Tabla 7.8** Optimización por el método simplex

	Factores					Respuesta
	A	B	C	D	E	
Vértice 1	1.0	3.0	2.0	6.0	5.0	7
Vértice 2	6.0	4.3	9.5	6.9	6.0	8
Vértice 3	2.5	11.5	9.5	6.9	6.0	10
Vértice 4 (rechazado)	2.5	4.3	3.5	6.9	6.0	6
Vértice 5	2.5	4.3	9.5	9.7	6.0	11
Vértice 6	2.5	4.3	9.5	6.9	9.6	9
(i) Suma (vértice 4 excluido)	14.50	27.40	40.00	36.40	32.60	
(ii) Suma /n (vértice 4 excluido)	2.90	5.48	8.00	7.28	6.52	
(iii) Vértice rechazado (es decir, 4)	2.50	4.30	3.50	6.90	6.00	
(iv) Desplazamientos = (ii) - (iii)	0.40	1.18	4.50	0.38	0.52	
(v) Vértice 7 = (ii) + (iv)	3.30	6.66	12.50	7.66	7.04	

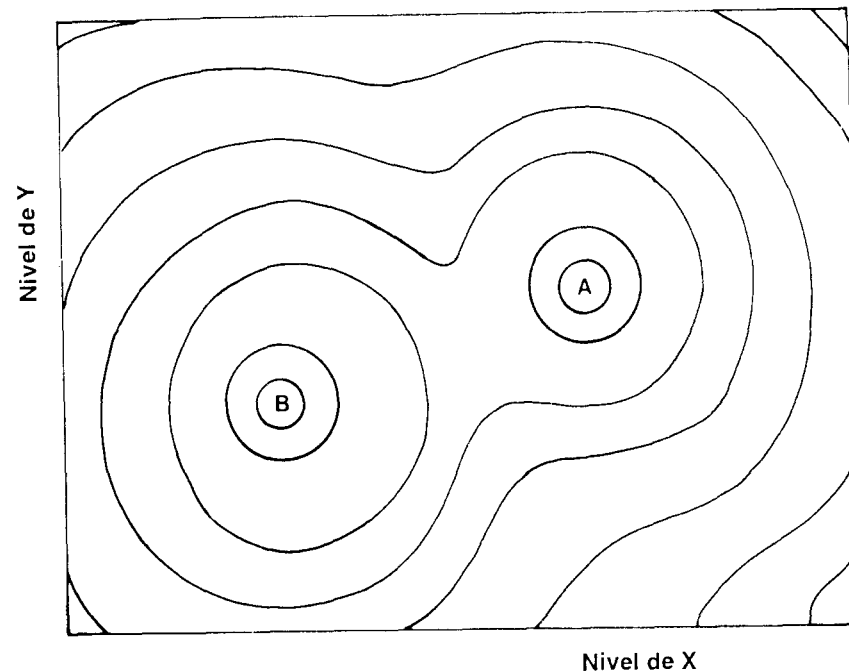
Deming (véase la bibliografía), que demuestran que el tamaño del simplex inicial no es tan crítico si se puede expandir o contraer según se procede con el método. Yarbro y Deming citan un algoritmo que se puede utilizar para calcular las posiciones iniciales de los vértices; un vértice normalmente se coloca en los niveles de los factores aceptados actualmente.

Se puede ver que, al contrario de un diseño factorial, el número de experimentos exigidos en el método simplex no crece rápidamente con el número de factores. Por esta razón todos los factores que se piense que pueden tener relación con la respuesta deberían ser incluidos en la optimización, Ya que al hacerlo así no se incrementará el número de experimentos necesario para definir el óptimo.

Una vez que se ha encontrado un óptimo, el efecto sobre la respuesta cuando se varía un factor mientras los otros se mantienen a sus niveles óptimos, se puede analizar cada vez para cada factor. Este procedimiento se puede utilizar para verificar la optimización. También indica la importancia de las desviaciones del nivel óptimo para cada factor; cuanto más acusado es el pico de respuesta en la región del óptimo, más crítica es la variación en el nivel del factor.

La optimización por simplex se ha utilizado con éxito en muchas áreas de química analítica, por ejemplo, espectrometría de absorción atómica, cromatografía de gases, métodos de análisis colorimétricos, espectrometría de plasma y analizadores centrifugos en química clínica. Cuando un instrumento se conecta con microcomputador, se pueden utilizar los resultados de la optimización del simplex para iniciar mejoras en las variables instrumentales.

Para superficies de respuesta con más de un pico, como la expuesta en la figura 7.11., los dos tipos de optimización descritos pueden encontrar un óptimo local tal como A en lugar del óptimo verdadero B. Una forma de verificar este punto es empezar en una región diferente y ver si se obtienen las mismas condiciones de óptimo. Los métodos de optimización proporcionan información sólo sobre algunas partes de la superficie de respuesta, ya que su objeto es maximizar la respuesta con el mínimo esfuerzo. Si se requiere un cuadro completo de la superficie de respuesta, son necesarios experimentos más elaborados que cubran todo el intervalo de las variables.



**Figura 7.11** Diagrama de líneas mostrando dos máximos localizados.

### 7.13 RECONOCIMIENTO DE PAUTAS

El avance en la automatización ha hecho posible la recolección rápida de grandes cantidades de datos, y el desarrollo de los microcomputadores ha hecho posible procesar estos datos con mayor perfeccionamiento. El término **quimiometría** se ha acuñado para describir los métodos estadísticos y matemáticos diseñados para extraer información química útil a partir de los datos químicos. La optimización es un campo cubierto por el paraguas de la quimiometría; otro es el campo del reconocimiento de pautas, que ahora pasamos a describir.

Una consecuencia de la automatización es que se pueden determinar muchos parámetros simultáneamente para la misma muestra, por ejemplo, en química clínica, cromatografía y análisis de emisión atómica. En química analítica se utilizan dichos datos en las pruebas de significación, por ejemplo, determinar si un derramamiento de aceite procede de una fuente concreta, utilizando las intensidades y/o las frecuencias de los muchos máximos obtenidos en el espectro de fluorescencia o infrarrojo de dichas muestras. Otro uso es la clasificación, por ejemplo, dividir las fases estacionarias utilizadas en cromatografía gas-líquido dentro de grupos con propiedades similares, utilizando las propiedades de retención de una variedad de solutos con propiedades químicas diferentes. En ambos casos los métodos de reconocimiento de pautas utilizan los datos disponibles simultáneamente en lugar de secuencialmente. El conjunto de medidas que se utiliza para caracterizar la muestra se denomina **pauta**. Cuando sólo se miden dos parámetros para cada muestra, la pauta se puede representar gráficamente por un punto, como se muestra en la figura 7.12,

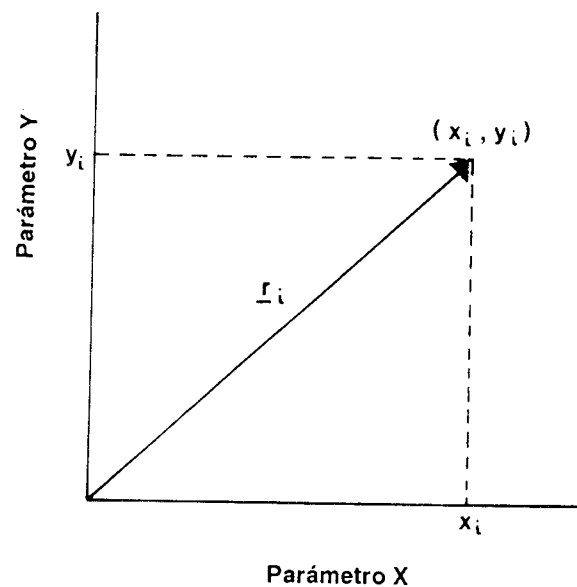


Figura 7.12 Una pauta representada en el espacio de pautas;  $x_i$  y  $y_i$  son los valores de los parámetros X y Y;  $r_i$  es el vector de pautas.

donde las coordenadas del punto son los valores tomados por los dos parámetros. El punto se puede definir también por un vector, dibujado desde el origen y conocido como un **vector de datos o de pautas**; el sistema coordenado se conoce como **espacio de pautas**. La base de todos los métodos de reconocimiento de pautas es que los vectores de pautas para muestras similares se sitúan con mucha proximidad en el espacio de pautas, formando **conglomerados**. En dos dimensiones este conglomerado se puede detectar con facilidad por el ojo humano. Sin embargo, cuando se miden más de dos parámetros, ya no es posible la representación gráfica: si se miden  $n$  parámetros, cada muestra se representará por un punto en un espacio  $n$  dimensional y se necesitan métodos matemáticos para detectar conglomerados. Uno de dichos métodos, conocido como **análisis de componentes principales**, permite proyectar los vectores de pautas sobre un plano, de tal forma que se pierda tan poca información como sea posible. La figura 7.13 ilustra el principio del método en la situación en que hay sólo dos parámetros. El diagrama muestra dos grupos bien definidos de vectores de pautas y dos líneas posibles sobre las que se podrían proyectar los puntos, reduciéndose de este modo el número de dimensiones de 2 a 1. La proyección sobre la línea A mantendría tanta diferencia como fuese posible entre los dos grupos, mientras la proyección sobre la línea B oscurecería la diferencia. En este caso, se debería hacer una proyección sobre la línea A. Las líneas A y B representan combinaciones lineales de las variables originales. De manera similar, el análisis de componentes principales permite hacer una proyección de  $n$  a 2 dimensiones, de tal forma que se conserva tanto como sea posible la diferencia entre los vectores de pautas. De nuevo, los ejes del plano sobre los que se realiza la proyección representan combinaciones lineales de las variables originales. En muchos casos, sólo algunas de las variables originales figuran significativamente en estas combinaciones lineales, y de este modo se pueden identificar las variables que

constituyen más a la variación entre los vectores de pautas. Las variables restantes, que llevan poca información extra, se pueden rechazar. De hecho, si se utilizan demasiadas variables, puede resultar una separación en clases aleatoria y sin sentido químico, o el "ruido" de los parámetros superfluos puede oscurecer la existencia de clases.

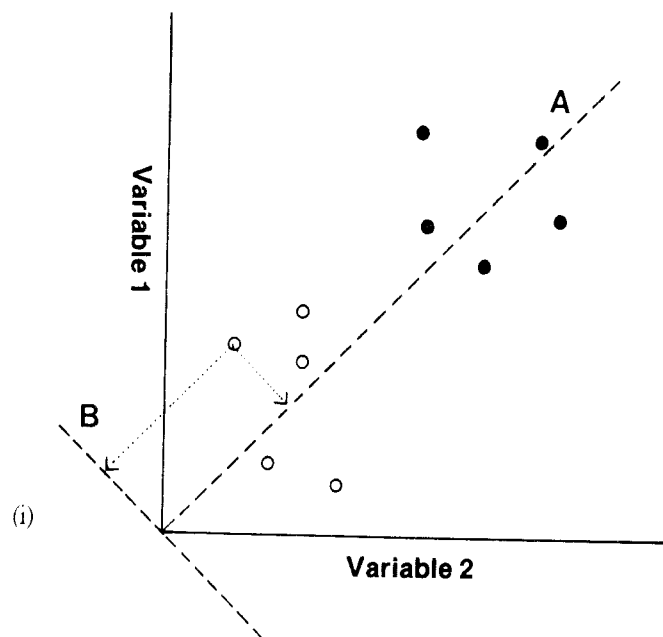
Esencialmente hay dos procedimientos diferentes para el reconocimiento de pautas, dependiendo si se utiliza para pruebas de significación o clasificación. En el primer caso se conocen las clases en que puede caer una muestra, y el objetivo es clasificar una muestra de clase desconocida por su pauta. En el caso del ejemplo de derramamiento de aceite, una muestra puede proceder de una de entre varias fuentes diferentes: cada fuente forma una clase y el propósito del reconocimiento de pautas es identificar la clase y, por lo tanto, la fuente de la muestra. En el segundo caso se trata de ver si las pautas obtenidas caen en grupos naturales; aquí no se espera ningún conocimiento *a priori* de las clases. Estas dos aproximaciones diferentes al reconocimiento de pautas se denominan, a veces, **aprendizaje supervisado** y **no supervisado**, respectivamente. Utilizan métodos diferentes y ahora pasamos a describirlas.

## 7.14 MÉTODOS DE APRENDIZAJE SUPERVISADOS

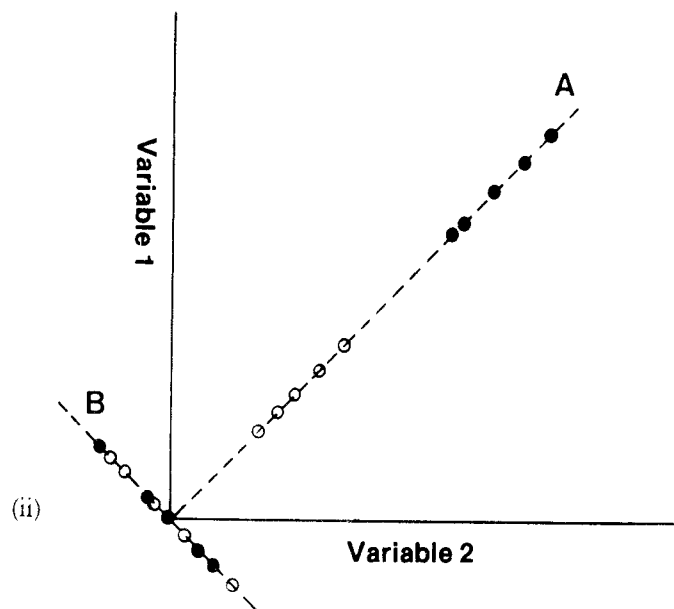
La situación más simple es aquella en la que una muestra puede pertenecer a una de dos clases: se conoce como **clasificación binaria**. Para el caso de dos parámetros se encuentra reflejada en la figura 7.14. Para encontrar un criterio o **clasificador** para separar las dos clases, se utiliza un conjunto de datos de pautas para muestras en las que se conoce la clase de la que son miembros. Este conjunto de datos se divide (utilizando números aleatorios) en dos partes llamadas el **conjunto de orientación** y el **conjunto de predicción**. El conjunto de orientación se utiliza para encontrar la posición de un plano, el **plano de decisión**, que separa las dos clases. Una manera simple de hacerlo es utilizar el plano de simetría entre los centros de gravedad de las dos clases. Esto significa que se asigna un vector de pautas desconocido a la clase que tenga su centro de gravedad más próximo al vector. La distancia  $d$  entre dos puntos en un espacio  $n$  dimensional con coordenadas  $(x_1, x_2, \dots, x_n)$  y  $(y_1, y_2, \dots, y_n)$  se calcula habitualmente como la distancia euclídea, dada por la expresión:

$$d^2 = \sum_{i=1}^n (x_i - y_i)^2$$

Este método es satisfactorio si las clases forman conglomerados compactos pero, si no es así, se puede obtener una separación más efectiva utilizando métodos más perfeccionados. Uno de estos es el **método de la máquina de aprendizaje**. Se trata de un procedimiento iterativo que utiliza retroalimentación (*feed-back*) negativa. Tomando en primer lugar como plano de decisión el plano de simetría entre los dos centros de gravedad, se prueban una a una las pautas del conjunto de orientación. Si se obtiene una clasificación incorrecta se ajusta el plano de decisión para rectificar esto. El método de la máquina de aprendizaje se denomina así debido a que "aprende" de los errores en la clasificación del reajuste requerido en la posición del plano de decisión. Si suponemos que se *puede* separar dos clases por un plano, este procedimiento conduce, eventualmente, a la posición de dicho plano. Si las clases *no* son separables por un plano, entonces es preferible el método de



(i)



(ii)

Figura 7.13 Proyección de puntos desde dos dimensiones a una dimensión; (i) muestra de puntos originales; (ii) muestra sus proyecciones sobre dos ejes diferentes.

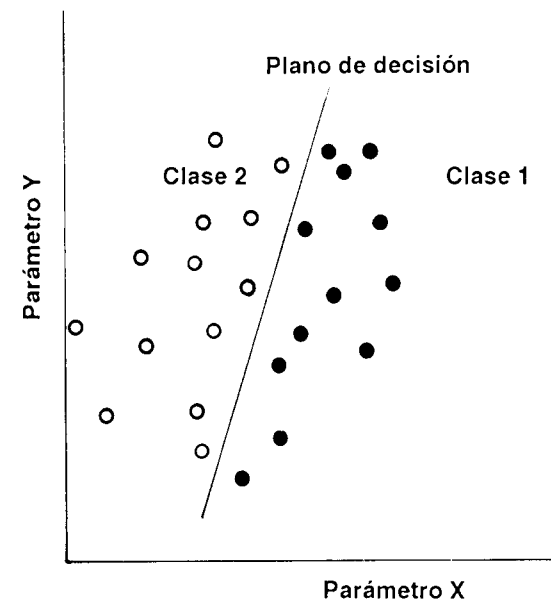


Figura 7.14 Dos clases separadas por un plano de decisión.

mínimos cuadrados. Este método minimiza una suma de cuadrados que estima los errores en la clasificación. Determinar la posición del plano de decisión se puede también tratar como un problema de optimización, resoluble por el método de simplex: el plano de decisión se puede definir por un vector ortogonal a él y las componentes de este vector son los factores que se variarán de nivel. La respuesta que se maximiza es la proporción de pautas del conjunto de orientación que se encuentran clasificadas correctamente.

Cuando la posición del plano de decisión se ha fijado, la efectividad del plano de decisión para separar las dos clases se puede probar utilizando el resto del conjunto de datos, es decir, el conjunto de predicción: cada muestra se clasifica por su pauta, y el porcentaje clasificado correctamente proporciona una estimación objetiva de la eficiencia del clasificador.

En la práctica, antes de aplicar el procedimiento de reconocimiento de pautas, se deberían escalar las variables (que pueden venir medidas en valores numéricos diferentes), de modo que no dominen las variables con los valores más grandes. Se puede hacer también un intento preliminar para ponderar las variables para optimizar la clasificación. Después de contrastar el clasificador, se pueden revisar de nuevo estos factores de ponderación.

Cuando hay más de dos clases que se van a separar, se puede utilizar una serie de clasificadores binarios. Un método alternativo, que es conceptualmente muy simple, es el **método del  $K$  entorno más próximo** (KNN, *K-nearest neighbour*, en inglés). Este método también tiene la ventaja de que se puede utilizar cuando no se pueden separar las clases por un plano, como se ilustra en la figura 7.15. En la forma más simple de separación, cada muestra se clasifica en la misma clase que el entorno más próximo. Alternativamente se puede tomar los  $K$  entornos más próximos (donde  $K$  es un entero pequeño) y decidir por



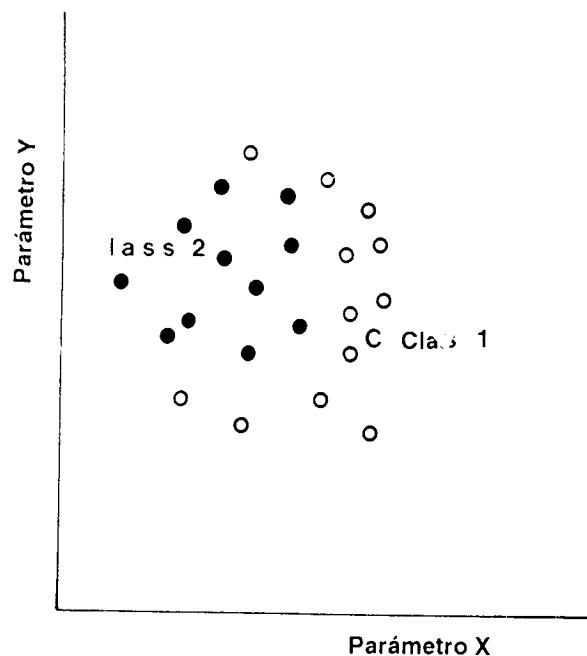


Figura 7.15 Dos clases que no se puede separar por un plano.

un esquema de votación los miembros de la clase. Se toma normalmente un valor impar para  $K$  a fin de evitar empates. En una versión más compleja se puede dar diferentes pesos a los entornos, dependiendo de su distancia relativa. Con este método no es necesario hacer ninguna división del conjunto de datos en conjuntos de predicción y orientación; se pueden utilizar todos los datos para estimar la eficiencia del método.

Las aproximaciones al reconocimiento de pautas descritas hasta ahora han considerado sólo la distancia entre puntos y no su orientación relativa en el espacio. Dichos métodos son no paramétricos, es decir, ningún modelo se propone para la distribución en el espacio de pautas de las muestras dentro de una clase. Por el contrario, el programa de computador SIMCA (debido a Wold), disponible comercialmente, utiliza el conjunto de datos para formar un modelo matemático de cada clase. Entonces es posible calcular para cada clase la probabilidad de que una muestra desconocida pertenezca a esa clase.

Algunos de los campos en que se han aplicado los métodos de reconocimiento de pautas supervisadas, son: pruebas de diagnóstico en química analítica, ciencia forense, la identificación de fuentes de polución, química de alimentos y la predicción de la estructura molecular a partir de los espectros.

## 7.15 MÉTODOS DE APRENDIZAJE NO SUPERVISADOS

El lector recordará que éstos son los métodos que se utilizan para decir si un conjunto de pautas se divide naturalmente en grupos. Existe un gran número de métodos de búsqueda de conglomerados en el espacio de pautas. Una aproximación agrupa, juntos, el par de

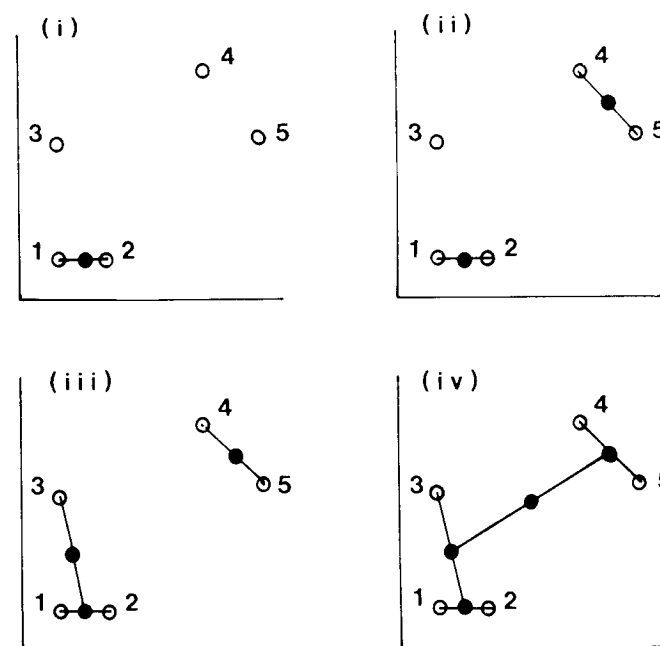


Figura 7.16 Etapas en el conglomerado de pautas: ○ representa un punto original, ● representa el único punto que se sustituye a un par de puntos.

puntos que están más próximos y los sustituye por un nuevo punto en la mitad de ambos. Este procedimiento se repite y si se continúa indefinidamente agrupará todos los puntos juntos como se muestra, por pasos, en la figura 7.16. En estos diagramas se unen por una línea los pares de puntos que han sido emparejados juntos y se sustituyen por un único punto. Así, en cada etapa, se unen los grupos de puntos que están clasificados juntos. Las etapas sucesivas del agrupamiento se puede mostrar en un **dendrograma** como el de la figura 7.17. El eje vertical muestra la **semejanza**,  $s_{ij}$ , entre los dos puntos  $i$  y  $j$ , se calcula según la expresión siguiente:

$$s_{ij} = 1 - d_{ij}/d_{\max}$$

donde  $d_{ij}$  es la diferencia entre los puntos  $i$  y  $j$  y  $d_{\max}$  es la distancia máxima entre cualquier par de puntos. Obviamente, la semejanza toma valores de 0 (para puntos coincidentes) a 1 (para los dos puntos más alejados). La etapa en la que se detiene el agrupamiento, y que determina el número de grupos en la clasificación final, se debería decidir por un criterio químico. Una aproximación opuesta trata todas las pautas inicialmente como un grupo, y después las subdivide.

Como se mencionó anteriormente, los métodos de aprendizaje no supervisado han sido utilizados para clasificar la gran cantidad de fases que se presentan en cromatografía gas-líquido. Se puede entonces seleccionar un pequeño conjunto referido de fases con características diferentes tomando una fase de cada clase. Otra aplicación es la clasificación de los antibióticos por su acción frente a los diferentes tipos de bacterias, para clasificar la relación entre actividad biológica y estructura molecular.

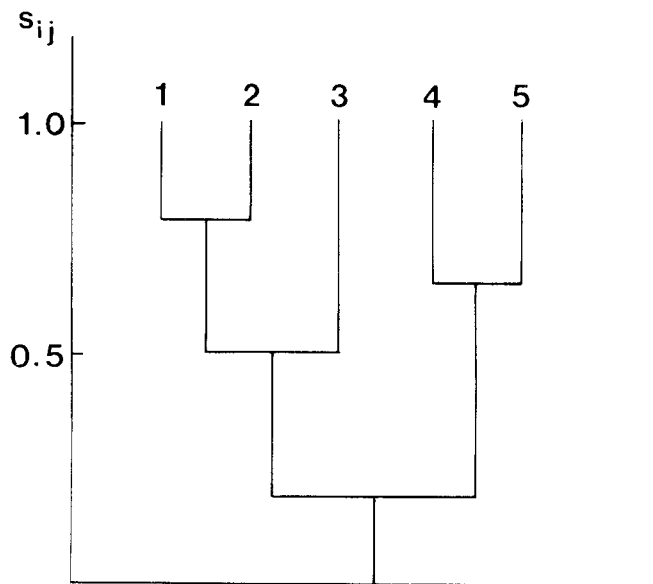


Figura 7.17 Un dendrograma para los puntos de la figura 7.16.

Queda claro que el reconocimiento de pautas se puede utilizar para establecer relaciones que de otra forma podrían quedar ocultas en una masa de datos. Es un método que se ha aplicado recientemente en química analítica: su potencial como un medio de análisis de datos no se ha evaluado por completo. Existen varios paquetes disponibles de programas de computador aparte del SIMCA. Uno de ellos es ARTHUR (desarrollado por Kowalski) que contiene programas para muchos métodos, incluidos los descritos aquí.

## BIBLIOGRAFÍA

Box, G. E. P., Hunter, W. G., y Hunter, J. S., *Statistics for Experimentalists*, Nueva York, Wiley, 1978. Proporciona un tratamiento detallado de los diseños factorial completo y fraccional y del método de la máxima pendiente.

Cooper, B. E., *Statistics for Experimentalists*, Oxford, Pergamon Press, 1975. Proporciona detalles adicionales de ANOVA.

Cox, D. R., *Planning of Experiments*, Nueva York, Wiley, 1958. Un tratamiento no matemático destinado al trabajador experimental. Proporciona detalles adicionales de los cuadrados Latinos.

Davis, J. C., *Statistics and Data Analysis in Geology*, Nueva York, Wiley, 1973. Informa con suficiente claridad del análisis de componentes principales.

Kowalski, B. R. (Ed.), *Chemometrics: Theory and Application*, Washington, American Chemical Society, 1977. Contiene artículos sobre SIMCA y ARTHUR.

Kryger, L., *Talanta*, 1981, **28**, 871. Un compendio de las aplicaciones de los métodos del reconocimiento de pautas a la química, incluyendo muchas referencias de trabajos originales.

Massart, D. L., Dijkstra, A., y Kaufman, L., *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*. Amsterdam, Elsevier, 1978. Una valiosa fuente de información de los temas de este capítulo. Incluye muchos ejemplos extraídos de la bibliografía y proporciona un gran número de referencias adicionales.

Varmuza, K., *Pattern Recognition in Chemistry*. Berlin, Springer-Verlag, 1980. Proporciona un tratamiento detallado de los métodos y aplicaciones del reconocimiento de pautas.

Yarbro, L. A., y Deming, S. N., *Anal. Chim. Acta*, 1974, **73**, 391. Discute la elección del simplex inicial.

Betteridge, D., Wade, A. P. y Howard, A. G., *Talanta*, 1985, **32**, 709, 723. Buena fuente de referencias de los métodos simplex, compara varias aproximaciones.

## EJERCICIOS

- En un experimento de colaboración entre laboratorios de la determinación de arsénico en carbón, se enviaron muestras de carbón de tres regiones diferentes a tres laboratorios. Cada laboratorio realizó un análisis por duplicado de cada muestra con los resultados que se muestran a continuación (medidos en  $\mu\text{g/g}$ ).

Muestra	Laboratorio		
	1	2	3
A	5.1, 5.1	5.3, 5.4	5.3, 5.1
B	5.8, 5.4	5.4, 5.9	5.2, 5.5
C	6.5, 6.1	6.6, 6.7	6.5, 6.4

Compruebe que no existe interacción significativa laboratorio-muestra y pruebe las diferencias significativas entre laboratorios.

- Si la respuesta en el vértice 7 en el ejemplo de la optimización simplex (pág. 170) es 12, ¿qué vértice se rechazará al formar un nuevo simplex y cuáles son las coordenadas del nuevo vértice?
- Se prepararon cuatro soluciones estándar, cada una conteniendo un 16% de cloruro calculado por pesada. Para analizar cada disolución estándar se emplearon tres métodos de titulación, cada uno con una forma diferente de determinar el punto final. El orden de los experimentos fue aleatorio. Los resultados del cloruro (%w/w) se muestran a continuación:

Solución	Método		
	A	B	C
1	16.03	16.13	16.09
2	16.05	16.13	16.15
3	16.02	15.94	16.12
4	16.12	15.97	16.10

Pruebe si existen diferencias significativas entre (a) la concentración de cloruro en las diferentes soluciones y (b) los resultados obtenidos por los diferentes métodos.

- Para conocer el contenido en nitrovina, se analizaron cuatro muestras diferentes de alimentos hervidos por dos métodos diferentes, cada uno por duplicado. Los resultados se muestran en la tabla siguiente (concentración en mg/kg). Pruebe si los resultados obtenidos por los métodos difieren significativamente.

Alimentos	Método HPLC	Método Espectrofotométrico
1	9.3, 10.0	9.0, 10.2
2	9.3, 12.3	9.0, 11.8
3	12.8, 12.7	12.9, 12.0
4	11.9, 11.8	11.6, 12.1

(Adaptado de Gliddon, M. J., Cordon, C. y Pamham, G. M., *Analyst*, 1983, **108**, 116.)

5. La temperatura efectiva del vapor de estaño, calculada de la razón de absorbancias de dos líneas espectrales, fue comparada con respecto a la temperatura real bajo diferentes condiciones experimentales en un tubo de cuarzo calentado. Los resultados se muestran en la tabla siguiente (valores en C). Calcule el efecto principal y de interacción.

		Temperatura efectiva			
		Gas de purga Ar		Gas de purga Ar + 1% O <sub>2</sub>	
Tiempo de purga (s)		20	60	20	60
Temperatura real (C)	900	843	826	852	855
	1000	908	873	908	908

(Adaptado de Welz, B. y Melcher, M., *Analyst*, 1983, 108, 213.)

6. De las soluciones guardadas en matraces de polipropileno el mercurio se pierde por combinación con el estaño adsorbido. Se midió la absorbancia de una solución acuosa estándar de mercurio, que se había guardado en dichos matraces, en dos niveles de los factores siguientes.

Factor	Bajo	Alto
A- agitación del matraz	Ausente	Presente
C- limpieza del matraz	Una vez	Dos veces
T- tiempo de permanencia	1 hora	18 horas

Se obtuvieron los resultados siguientes: calcule el efecto principal y de interacción.

Combinación de los niveles de factores	Absorbancia
l	0.009
a	0.084
c	0.097
t	0.076
ac	0.082
ta	0.049
tc	0.080
atc	0.051

(Adaptado de Kuldvere, A., *Analyst*, 1982, 107, 179.)

7. Se investigó el porcentaje de recuperación de plomo añadido a muestras en presencia de concentraciones “bajas” y “altas” de ácido nítrico y lantano por espectrometría de absorción atómica electrotrémica; a continuación se proporcionan las medias de análisis duplicados. Calcule el efecto principal y de interacción. Si el cuadrado medio residual es 4 (con 4 grados de libertad), pruebe si el efecto de interacción es significativo.

Concentración de lantano (g/100 ml)		
Concentración de ácido nítrico (%v/v)	0	0.10
	64	90
	77	96

(Adaptado de Bertenshaw, M. P., Gelsthorpe, D. y Wheatstone, K. C., *Analyst*, 1982, 107, 163.)

# Soluciones a los ejercicios

## CAPÍTULO 1

- El laboratorio A ha conseguido un resultado medio de 41.9 g/l, (muy cercano al valor correcto) y una pequeña variabilidad en los resultados —todos los valores caen entre 41.1 y 42.5 g/l. Estos resultados son exactos y precisos: los errores sistemáticos y aleatorios son pequeños. El laboratorio B ha conseguido el mismo valor medio, pero la variabilidad de los resultados es mucho mayor (intervalo 39.8-43.9 g/l), de manera que, aunque no existen aparentemente errores sistemáticos (es decir, el resultado medio es exacto), existen grandes errores aleatorios (es decir, los datos son muy imprecisos). Consideraciones similares muestran que el laboratorio C ha obtenido resultados precisos pero inexactos (media 43.2 g/l), y que los resultados del laboratorio D son imprecisos e inexactos. El laboratorio E ha obtenido una serie de resultados que parecen precisos y exactos, con la excepción del valor final. En la práctica, se podría probar que éste es un valor anómalo (veáse el Cap. 3): si las pruebas demostrasen que se pudiera rechazar, con un razonable nivel de confianza, los resultados restantes serían muy similares a los del laboratorio A.
- La segunda serie de seis resultados obtenida por el laboratorio A tiene el mismo resultado medio que la primera serie, confirmando que este laboratorio obtiene resultados con errores sistemáticos no significativos. En la segunda serie la variabilidad es más grande (intervalo 40.8-43.3 g/l), ilustrando la diferencia entre la precisión dentro de los días o dentro del lote (repetitividad) y la precisión entre días o entre lotes (reproducibilidad).
- Las poblaciones de anticuerpos monoclonales son homogéneas, de aquí que el número de posiciones de enlace por molécula deba ser un número entero —obviamente dos en este caso. Los resultados son, en este caso, precisos, pero muestran clara evidencia de un error sistemático que produce valores bajos.
- (i) La concentración de lactato en sangre humana varía ampliamente de un paciente sano a otro (ca. 5-20 mg/100 ml), y también varía con menor extensión dentro del

mismo sujeto de una vez a otra. Si se estudia el grado del último tipo de variación, la exactitud no será muy importante pero serán necesarias mediciones precisas —los errores experimentales deben ser despreciables comparados con las variaciones del individuo. Si sólo se realiza una medición, para verificar si un sujeto cae o no cae dentro del “intervalo normal” de lactato en sangre, se requerirá menos precisión, aunque un gran error sistemático podría conducir a un diagnóstico equivocado.

(ii) El contenido de uranio en minerales se estudia con vistas a una extracción del elemento económicamente viable; por lo tanto no será necesaria una alta precisión, ¡sin embargo, errores sistemáticos sustanciales (positivos o negativos) podrían conducir a decisiones económicamente desastrosas!

(iii) En este análisis la velocidad es claramente esencial, y ni la exactitud ni la precisión son inicialmente de gran importancia. A medida que el paciente envenenado se recupera, el nivel en plasma de la droga puede ser repetidamente controlado para confirmar que está decreciendo: ya que se estudia una tendencia, la precisión es más importante que la exactitud.

(iv) Aquí, de nuevo, el objetivo principal es detectar cambios en el resultado del análisis; como estos cambios pueden ser muy pequeños, será necesaria buena precisión para detectar cualquier tendencia, mientras que la exactitud será de menor importancia.

5. (i) En este experimento, la fuente más probable de error es que la muestra tomada no sea representativa del conjunto mineral, y de este modo se puede dar un valor erróneo del contenido de Fe a granel (el muestreo se analizó en el capítulo 4). Puede surgir también un error sistemático si la reducción de Fe(III) a Fe(II) es incompleta, o si existe un gran error de indicador. Otros errores sistemáticos, aparte del error de muestreo, se pueden detectar con la ayuda de una muestra de mineral estándar —tales muestras están disponibles comercialmente y tienen un valor certificado de Fe. Otro problema, no siempre resuelto utilizando la muestra estándar, es la posibilidad de que otros elementos en estados de oxidación bajos serán titulados por el sulfato cérico, dando un resultado falso de hierro por exceso. Los errores aleatorios en análisis volumétrico se estudian en las secciones preliminares de este capítulo.

(ii) Aparte de los errores sistemáticos vistos en (i), una formación y/o extracción incompleta del quelato plantearán, en este caso, los mayores problemas. De nuevo, tales errores se detectarán con la ayuda de una muestra mineral de contenido de Fe conocido. El análisis colorimétrico utilizará probablemente una serie de patrones de Fe y un procedimiento gráfico para calcular el resultado y los errores aleatorios: dichos métodos se analizan posteriormente en el capítulo 5.

(iii) Los errores aleatorios en análisis gravimétrico deberían ser muy pequeños y los errores sistemáticos, directamente asociados con el proceso de pesada pueden minimizarse mediante una técnica cuidadosa (véase la Sec. 1.4). Las fuentes de errores más probables en esta determinación son químicas, por ejemplo el problema de coprecipitación de otros iones, y se discuten ampliamente en libros de texto de análisis clásico.

## CAPÍTULO 2

1. Media =  $0.077 \mu\text{g/g}$ . Desviación estándar =  $0.007 \mu\text{g/g}$ . DER = 9%.

2. Media = 5.163. Desviación estándar = 0.027. Número de grados de libertad = 6; (i)  $5.16_3 \pm 0.02_5$ ; (ii)  $5.16_3 \pm 0.03_8$ .
3. Número de grados de libertad = 9; los límites de confianza al 95% son  $20.9_2 \pm 0.3_4 \text{ g/l}$ .
4. Número de grados de libertad = 49; los límites de confianza al 95% son  $10.1_2 \pm 0.1_8 \text{ ng/ml}$ . Para reducir la longitud del intervalo de confianza se requiere un tamaño de muestra más grande y es seguro suponer que  $t$  estará cercano a 1.96. El tamaño de la muestra se calcula de  $0.1 = 1.96 \times 0.64/\sqrt{n}$ , dando  $n = 157$ .
5. Media = 10.45%. Desviación estándar = 0.105%. Número de grados de libertad = 5. Los límites de confianza al 95% son  $10.4_5 \pm 0.1_1$ . Los límites de confianza al 99% son  $10.4_5 \pm 0.1_7\%$ . (i) No; (ii) no. Esto sugiere que el método da un error sistemático.
6. La media y la desviación estándar son 1.74 y 1.28 g/l, respectivamente. Los logaritmos decimales de los valores dados son 0.260, 0.521, 0.029, 0.104, -0.310, 0.579, -0.824, 0.297. La media de estos valores es 0.082 y su desviación estándar es 0.462. Número de grados de libertad = 7. El intervalo de confianza al 95% para los valores logarítmicos es  $0.082 \pm 0.386$ , es decir, de -0.304 a 0.468. Tomando antilogaritmos da como intervalo de confianza de la media 0.50 a 2.94. Observe que este intervalo no es del todo simétrico alrededor de la media.
7. Media = 0.2943. Desviación estándar = 0.00468. Número de grados de libertad = 9. Los límites de confianza al 99% son  $0.294_3 \pm 0.004_8$ .
8. Media = 10.19. Desviación estándar = 0.19. Número de grados de libertad = 4. Los límites de confianza al 95% son  $10.1_9 \pm 0.2_3$ , es decir, 9.96 - 10.42. Ya que el intervalo de confianza incluye 10.00 no existe evidencia de error sistemático.
9. Desviación estándar = 0.14 mg. La pesada del reactivo requerido es 0.5 (4.9) g dando una DER = 0.028 (0.0029)%. DER para el volumen de disolvente = 0.02%. DER de la molaridad de la solución = 0.034 (0.020)%. Los valores para el reactivo con peso-fórmula 392 se dan entre paréntesis.
10. Solubilidad =  $\sqrt{(1.3 \times 10^{-10})} = 1.14 \times 10^{-5} M$ . DER del producto de solubilidad = 7.7%, dando DER de la solubilidad = 3.85% y desviación estándar =  $0.04_4 \times 10^{-5} M$ .

## CAPÍTULO 3

1. Los puntos caen aproximadamente en una línea recta, indicando que los datos se han extraído de una distribución normal.
2.  $Q = 0.13/0.23 = 0.565$ . Valor crítico = 0.570. El resultado no es significativo al nivel de confianza  $P = 0.05$ ; no existe evidencia que 2.07 sea un valor anómalo.
3. Tomato: media = 772.6, desviación estándar =  $13.6 \mu\text{g/g}$ . Pepino: media = 708.9, desviación estándar =  $10.4 \mu\text{g/g}$ . (i) Se requiere un prueba de dos colas.  $F_{6,6} = 1.70$ .

Valor crítico = 5.82 ( $P = 0.05$ ). Las varianzas no difieren significativamente. (ii) Ya que las varianzas no difieren significativamente, se calcula una estimación conjunta de la desviación estándar y es 12.1  $\mu\text{g/g}$ ;  $t = 1.28$ . Para una prueba de dos colas, el valor crítico = 2.18 ( $P = 0.05$ ), de manera que las medias no difieren significativamente.

4. Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrados medios
Entre muestras	6365.7116	3	2121.9
Dentro de muestras	162.0284	20	8.1
	6527.74	23	

Para probar si el porcentaje de agua recogida a diferentes profundidades difiere significativamente, se compara la variación entre muestras con la variación dentro de muestras:  $F_{3,20} = 262$ . Valor crítico = 3.098 ( $P = 0.05$ , prueba de una cola). El valor es significativo, indicando que hay diferencia entre las medias muestrales. Las medias muestrales son 34.1, 45.4, 72.2, 70.4. La mínima diferencia significativa = 3.43, calculada tomando  $t_{20} = 2.09$  ( $P = 0.05$ , prueba de dos colas). Todos los pares de profundidades, excepto el par más profundo, difieren significativamente unos de otros.

5. Use la prueba chi-cuadrada con cada frecuencia esperada = 5.  $\chi^2 = 16.8$ . Hay 9 grados de libertad y el valor crítico es 16.92 ( $P = 0.05$ ). No existe evidencia suficiente en la preferencia de algunos dígitos.

6.  $t = 1.54, 1.60, 1.18, 1.60$ . Hay 7 grados de libertad y el valor crítico es 2.36 ( $P = 0.05$ , prueba de dos colas). Ninguna de las medias difiere significativamente del valor certificado.

7. Las medias son 55.875 y 50.25 mg/kg y las desviaciones estándar son 0.8345 y 8.828 mg/kg.

(a)  $F_{7,7} = 112$ . Valor crítico = 3.79 ( $P = 0.05$ , prueba de una cola). Variabilidad significativamente más grande para tiempos de ebullición más largos.

(b) Una estimación conjunta de la varianza no es adecuada cuando las varianzas difieren significativamente;  $t = 1.8$ . El número de grados de libertad, calculado de la ecuación (3.5) es 7. Valor crítico = 2.36 ( $P = 0.05$ , dos colas). Las medias no difieren significativamente.

8. (a) Frecuencias esperadas = 15.25 y 45.75, suponiendo que el número de roturas está en proporción al número de trabajadores en cada grupo, mientras que las frecuencias reales son 24 y 37; 1 grado de libertad  $\chi^2 = 5.95$  (Se aplica la corrección de Yates). Valor crítico = 3.84 ( $P = 0.05$ ). El primer trabajador difiere significativamente de los otros tres.

(b) Para los últimos tres trabajadores las frecuencias esperadas = 12.33, 12.33, 12.33 (es decir, 37/3);  $\chi^2 = 2.81$ ; 2 grados de libertad. Valor crítico = 5.99 ( $P = 0.05$ ). Los últimos tres trabajadores no difieren significativamente unos de otros.

9. En primer lugar debemos probar si las varianzas difieren significativamente:  $F_{2,2} = 1.22$  —no significativo ( $P = 0.05$ , dos colas). Entonces se puede calcular una estima-

ción conjunta de la varianza que es igual a 0.1685;  $t = 0.54$ . Valor crítico = 2.78 ( $P = 0.05$ , dos colas). Los resultados medios no difieren significativamente.

10. En este ejemplo una prueba  $t$  por parejas es adecuada cuando se aplican ambos métodos a cada muestra: pruebe que las diferencias entre pares elimina el efecto de cualquier variación entre las muestras. Las diferencias son 0.051, 0.030, -0.030 mg/ml. Estas diferencias tienen una media = 0.017 mg/ml, desviación estándar = 0.042 mg/ml. Hay 2 grados de libertad y  $t = 0.7$ . Valor crítico = 4.3 ( $P = 0.05$ , dos colas). Los resultados no difieren significativamente.

11. Hombres: media = 40.0 g/l, desviación estándar = 2.78 g/l. Mujeres: media = 43.25 g/l, desviación estándar = 3.06 g/l. En primer lugar debemos probar si las varianzas difieren significativamente:  $F_{7,7} = 1.21$ . El valor crítico es 4.995 ( $P = 0.05$ , dos colas), de manera que las varianzas no difieren significativamente. La estimación conjunta de la desviación estándar es 2.92;  $t = 2.2$ . Valor crítico = 2.14 ( $P = 0.05$ , dos colas). Las medias difieren significativamente.

12. Una prueba  $t$  por parejas es adecuada. La media y la desviación estándar de las diferencias son 1.18 y 0.84 respectivamente;  $t = 3.4$ . Valor crítico = 2.57 ( $P = 0.05$ , dos colas). Los resultados difieren significativamente.

## CAPÍTULO 4

1. Esquema 1:  $\sigma^2 = (4/2) + (10/5) = 4$

Esquema 2:  $\sigma^2 = (4/2 \cdot 3) + (10/3) = 4$

Si  $S$  = costo de muestreo y  $A$  = costo del análisis, entonces:

costo del esquema 1: costo del esquema 2 =  $(5S + 2A) : (3S + 6A)$ .

Esta razón es mayor que 1 si  $S:A$  es mayor que 2:1.

2. Los cálculos del ANOVA se resumen a continuación:

Fuente de variación	Suma de cuadrados	g.l.	cuadrado medio
Entre días	332.92	3	111
Dentro de días	26.00	8	3.25
Total	358.92	11	

$F_{3,8} = 34$ , Valor crítico = 4.066 ( $P = 0.05$ ) demostrando que las concentraciones medias difieren significativamente. Estimación de la varianza muestral =  $(111 - 3.25)/3 = 35.9$ .

3. Los cálculos del ANOVA se resumen a continuación:

Fuente de variación	Suma de cuadrados	g.l.	cuadrado medio
Entre muestras	0.0024916667	3	0.0008306
Dentro de muestras	0.0014	8	0.000175
Total	0.0038916667	11	

$F_{3,8} = 4.746$ . El valor crítico ( $P = 0.05$ ) es 4.066, demostrando que el cuadrado medio entre muestras es demasiado grande para ser explicado por variación en las mediciones.

El cuadrado medio dentro de muestras, 0.000175, proporciona una estimación de la varianza de las mediciones,  $\sigma_2^2$ .

Estimación de la varianza muestral,  $\sigma_1^2 = 0.0008306 - 0.000175/3 = 0.000219$ .

Varianza de la media para el esquema 1 =  $0.000175/4 + 0.000219/6 = 0.00008025$ .

Varianza de la media para el esquema 2 =  $[0.000175/(2 \times 3)] + 0.000219/3 = 0.0001022$ .

4. Se realiza una estimación de  $\sigma^2$  de cada muestra, dando: 2.607, 0.847, 3.017, 3.633, 6.417, 1.927. El promedio de estos valores es 3.077. Entonces la estimación de  $\sigma = \sqrt{3.077} = 1.753$ .

Las líneas de alerta se encuentran en  $50 \pm 2 \cdot 1.753/\sqrt{4}$ , es decir, en 48.25 y 51.75

Las líneas de acción se encuentran en  $50 \pm 3 \cdot 1.753/\sqrt{4}$ , es decir, en 47.37 y 52.63.

5. Cálculos simples muestran que las concentraciones medias de Cd en las muestras A y B, determinadas por los 15 laboratorios, son 7.01 y 7.75 ppm respectivamente. Cuando se representa en una carta de dos muestras el punto (7.75, 7.01), y se dibujan a través de él la línea horizontal y vertical, los cuatro cuadrantes generados contienen (leyendo en el sentido de las agujas del reloj a partir del cuadrante +, +) 6, 1, 7 y 1 puntos. Este resultado es típico de un ejercicio de colaboración de este tipo, y sugiere que las diferencias grandes entre los 15 laboratorios son debidas fundamentalmente a errores sistemáticos. Hay que hacer constar que sólo cuatro laboratorios (7, 9, 12 y 13) generan resultados que están bastante cercanos a las medias de ambas muestras, mientras que los laboratorios 3 y 10 poseen para ambas muestras grandes sesgos positivos y negativos, respectivamente. Las distancias en perpendicular desde los puntos experimentales a la línea de 45° que pasa por (7.75, 7.01) mide los errores aleatorios de los laboratorios participantes, y queda claro que los laboratorios 4 y 14 pueden tener grandes errores aleatorios en sus resultados.

Para calcular los componentes de la variación, se expone la tabla siguiente:

Laboratorio	A	B	D	T
1	8.8	10.0	-1.2	18.8
2	3.8	4.7	-0.9	8.5
3	10.1	12.1	-2.0	22.2
4	8.0	11.0	-3.0	19.0
5	5.0	4.7	0.3	9.7
6	5.2	6.4	-1.2	11.6
7	6.7	8.4	-2.0	15.4
8	9.3	9.6	-0.3	18.9
9	6.9	7.5	-0.6	14.4
10	3.2	2.8	0.4	6.0
11	9.7	10.4	-0.7	20.1
12	7.2	8.3	-1.1	15.5
13	6.5	6.8	-0.3	13.3
14	9.7	7.2	2.5	16.9
15	5.0	6.0	-1.0	11.0

De las columnas D y T de la tabla, es fácil calcular que la estimación de  $\sigma_0^2 = 0.793$ , y que la estimación de  $\sigma^2 = 11.027$ . El valor experimental de  $F_{14,14}$  es 11.027/0.793 = 13.095: éste es mucho mayor que el valor tabulado (ca. 2.48, obtenido de la tabla A.2 por interpolación), mostrando que los errores sistemáticos son de hecho significativos. Se pueden estimar de la ecuación (4.3), siendo el resultado de  $\sigma_1^2 = 5.117$ .

## CAPÍTULO 5

1. La aplicación de la ecuación (5.2) a los datos proporciona un coeficiente de correlación producto-momento,  $r$ , de -0.8569. Este valor sugiere una correlación negativa considerable entre el nivel de mercurio y la distancia al polarógrafo. Su significación puede ser probada utilizando la ecuación (5.3), que proporciona un valor de  $t$  de 3.33. La comparación de este valor con el valor tabulado de  $t$  a un nivel de confianza del 95% y  $n - 2$  (= 4) grados de libertad ( $t = 2.78$ ) muestra que la relación es de hecho significativa a este nivel de confianza. Sin embargo, hay dos advertencias, (1) este método prueba sólo una correlación lineal, mientras que en el laboratorio podríamos esperar que el nivel de mercurio decreciera con el incremento de la distancia al polarógrafo de forma inversamente proporcional al cuadrado de la misma; (2) correlación *no* es lo mismo que causa. Los cálculos muestran que existe una relación entre la distancia al polarógrafo y la concentración de mercurio —*no* prueban que el polarógrafo sea la fuente de la contaminación de mercurio. La contaminación podría surgir, por ejemplo del fregadero o de otra pieza del equipo próxima al polarógrafo.
2. En este caso la ecuación (5.2) muestra que  $r = 0.99982$ . Esto es sintomático de una correlación lineal, *aunque* inspeccionando los datos se muestra claramente que una curva es en más adecuada en este caso. Cada valor de  $y$  (absorbancia) es más grande que su predecesor, pero con un incremento que decrece lentamente cuando  $x$  aumenta. La curvatura también puede demostrarse estudiando los residuos (véase el ejercicio 9). En este caso, los errores causados al tratar los datos como en una gráfica de calibración lineal serían muy pequeños.
3. Aplicando las ecuaciones (5.4) y (5.5) resultan unos valores de  $b = 0.0252$ , y  $a = 0.0021$  respectivamente. Otros resultados importantes incluyen:  $\bar{x} = 15$ ;  $\bar{y} = 0.380$ ;  $\sum x_i^2 = 2275$ ; y  $\sum (x_i - \bar{x})^2 = 700$ . Los residuos individuales ( $y_i - \hat{y}_i$ ) de  $y$  encontrados son +0.0009, -0.0009, -0.0028, +0.0104, -0.0074, -0.0062, +0.0060. Estos residuos suman cero (como era de esperar), y la suma de sus cuadrados es 0.000247. Aplicando la ecuación (5.6) se obtiene que  $s_{yx} = 0.00703$ . Entonces la ecuación (5.7) da  $s_b = 0.000266$ , y el uso de  $t$  (5 grados de libertad; nivel de confianza del 95%) = 2.57 proporciona un límite de confianza para  $b$  de  $0.0252 \pm 0.0007$ . Similarmente, la ecuación (5.8) proporciona un valor de  $s_a = 0.00479$ , con límites de confianza para  $a$  de  $0.0021 \pm 0.0123$ .
4. (a) Es fácil demostrar, usando los valores de  $a$  y  $b$  calculados en el ejercicio 3, que un valor de  $y$  de 0.456 corresponde a un valor de  $x$  de 18.04 ng/ml, y que  $(y_0 - \bar{y})^2 = 0.00578$ . La aplicación de la ecuación (5.9) proporciona entonces un valor de  $s_{x0} = 0.300$ , y los correspondientes límites de confianza están dados por  $18.04 \pm (2.57 \times 0.300) = 18.04 \pm 0.77$  ng/ml.

(b) En este caso, es claramente necesario ver si el valor de absorbancia 0.347 es un valor anómalo. El valor de  $Q$  calculado de la ecuación (3.8) es  $(0.347 - 0.314)/(0.347 - 0.308) = 0.846$ . Para un nivel de confianza del 95%, el valor obtenido es mayor que el valor de  $Q$  tabulado de 0.829, de manera que el valor 0.347 *puede* rechazarse. De los tres resultados que quedan se obtiene una absorbancia media de 0.311, correspondiente a una concentración de 12.28 ng/ml. En este caso el valor de  $s_{x0}$  está dado por la ecuación (5.10) [ $m = 3$ ;  $(y_0 - \bar{y})^2 = 0.00476$ ] siendo 0.195, los límites de confianza son, por tanto,  $12.28 \pm (2.57 \times 0.195) = 12.28 \pm 0.50$  ng/ml.

- Con los resultados del anterior ejercicio 3, este cálculo es muy sencillo. El límite de detección se define como la concentración que produce una señal que supera a la señal de absorbancia de fondo en tres veces la desviación estándar, la señal de fondo viene dada por  $a$  (0.0021) y la desviación estándar por  $s_{y/x}$  (0.00703). La señal de absorbancia en el límite de detección es  $0.00211 + (3 \times 0.0073) = 0.0232$ , y el límite de detección, calculado de la pendiente y ordenada en el origen de la recta de regresión, es 0.84 ng/ml.
- Los cálculos de regresión lineal usuales dan  $a = 0.2569$  y  $b = 0.005349$ . La razón de estos dos valores da la concentración de la muestra problema (48.0 ng/ml). También  $\bar{x} = 35$ ,  $\bar{y} = 0.4441$ ,  $\sum(x_i - \bar{x})^2 = 4200$ , y  $\sum(y_i - \hat{y}_i)^2 = 0.00001364$ . Así pues  $s_{y/x} = 0.003693$ , y  $s_{x0}$  dado por la ecuación (5.12) es 0.9179. Los límites de confianza para la concentración son  $48.03 \pm (2.45 \times 0.9179) = 48.0 \pm 2.2$  ng/ml.
- Los datos disponibles permiten un cálculo de regresión ponderado, siendo los valores de  $y$ , y desviaciones estándar:

$x$ :	0	10	20	30	40	50
$y$ :	4.0	21.2	44.6	61.8	78.0	105.2
$s$ :	$0.7_1$	0.8 <sub>4</sub>	0.8 <sub>9</sub>	1.6 <sub>4</sub>	2.2 <sub>4</sub>	3.0 <sub>3</sub>

El primer paso es calcular la recta de regresión no ponderada siendo la pendiente y la ordenada en el origen 1.982 y 2.924 respectivamente. Las concentraciones correspondientes a intensidades de fluorescencia de 15 y 90 unidades son de este modo 6.09 y 43.9 ng/ml respectivamente. De la manera habitual (véase el ejercicio 3),  $s_{y/x} = 2.991$ . Ya que se calcula fácilmente que  $\sum(x_i - \bar{x})^2$  es 1750 y  $\bar{y}$  es 52.47, los valores de  $s_{x0}$  para las dos concentraciones son en los dos casos 1.767 [Ec. (5.9)]. Usando un valor de  $t$  de 2.78, los límites de confianza al 95% para estos valores están dados por  $6.1 \pm 4.9$  y  $43.9 \pm 4.9$  ng/ml respectivamente.

Para calcular la recta de regresión ponderada, es necesario, en primer lugar, calcular la ponderación de cada punto utilizando la ecuación (5.13). A fin de incrementar los valores de  $x_i$ , estas ponderaciones son 2.227, 1.591, 1.418, 0.417, 0.224 y 0.122 —nótese que la suma de estos números es 5.999, es decir 6, si redondeamos errores. Ésta es una importante verificación por realizar. Utilizando estas ponderaciones y las ecuaciones (5.14) y (5.15) la pendiente y la ordenada en el origen de la recta de regresión ponderada son 1.964 y 3.483 respectivamente. De esta recta, las concentraciones correspondientes a las intensidades de fluorescencia de 15 y 90 unidades son

5.87 y 44.1 ng/ml respectivamente. Note que estos valores, y los de la pendiente y ordenada en el origen, son todos similares a los calculados en la regresión no ponderada, a pesar de la considerable dispersión de los puntos experimentales. La diferencia esencial entre el uso de los métodos de regresión ponderada y no ponderada sólo es patente cuando la ecuación (5.16) se utiliza para estimar la desviación estándar de las concentraciones estimadas. Para estimar esta ecuación, es necesario, en primer lugar, estimar los valores  $w_0$  para los dos valores  $y_0$  bajo estudio. La inspección de los datos sugiere que pesos de 1.8 y 0.18 respectivamente son apropiados para los valores de  $y_0$  de 15 y 90. El uso de estos pesos da valores de  $s_{x0w}$  de 0.906 y 2.716 respectivamente. La multiplicación por el valor de  $t$  de 2.78 proporciona límites de confianza para las dos concentraciones de  $5.9 \pm 2.5$  y  $44.0 \pm 7.6$  ng/ml respectivamente.

- Aquí, el método “nuevo” es claramente el ESI, representado, por lo tanto, en el eje  $y$ : la técnica gravimétrica es la referencia o procedimiento estándar y se representa en el eje  $x$ . La recta de regresión calculada de esta manera tiene una pendiente de 0.963, y una ordenada en el origen de 4.48, siendo el coeficiente de correlación de 0.970. Este último valor es sumamente significativo [ $t = 11.29$  - Ec. (5.3)], pero los límites de confianza de la pendiente y la ordenada en el origen deben calcularse también. Los cálculos usuales [Ecs. (5.6)-(5.8)] y un valor tabulado de  $t = 2.31$  (8 grados de libertad; a un nivel de confianza del 95%) muestran que los límites de confianza para la pendiente y la ordenada en el origen son respectivamente  $0.96 \pm 0.200$  y  $4.5 \pm 20.1$ . Ya que estos intervalos de confianza al 95% incluyen claramente a 1 y 0 respectivamente, debemos concluir que esta recta de regresión indica una buena concordancia entre los dos métodos analíticos. No obstante, hay una sensación residual de intranquilidad sobre los datos. Los cálculos muestran que 7 de los residuos y son negativos (incluyendo 6 en una secuencia simple) y sólo 3 son positivos. Se sugiere también que la concordancia entre los dos métodos es muy grande para la mayoría de las muestras, pero muy pobre para una pocas; esto podría ser causado por problemas de especiación. Un buen analista podría concluir que se necesitan más datos (en el trabajo citado se suministran más datos).

- Si se representan gráficamente estos datos (recuerde que debería hacerse *siempre*), existen ciertos indicios de que la calibración sea lineal hasta 0.7 - 0.8 unidades de absorbancia, y curva posteriormente. Cálculos de regresión lineal no ponderada utilizando los seis puntos proporcionan un coeficiente de correlación de 0.9936, y los residuos  $y$  (en orden creciente de  $x$ ) de -0.07, -0.02, +0.02, +0.06, +0.07 y -0.07 (redondeando a dos cifras decimales). La tendencia en estos valores es obvia y confirma que se trata de una curva. La suma de los cuadrados de los residuos es 0.0191. La confirmación de que el último punto debería ser omitido del intervalo lineal, se obtiene repitiendo el cálculo de regresión sólo con los cinco puntos. El coeficiente de correlación entonces aumenta a 0.9972 y los residuos  $y$  son -0.04, 0, +0.02, +0.04 y -0.02 respectivamente (suma de cuadrados 0.0040). Los residuos  $y$  para el sexto punto con respecto a esta segunda línea recta es comparativamente muy grande: -0.22.

Este proceso puede repetirse para el quinto punto, es decir, se calcula la línea recta ajustada sólo con los cuatro primeros puntos. De nuevo se mejora, en este caso, el coeficiente de correlación (a 0.9980), se reducen los valores de los residuos y (-0.02,

+0.01, +0.01 y -0.01), y existe un gran residuo para el quinto punto bajo estudio (-0.14). Puede concluirse, por lo tanto, que el quinto punto puede excluirse también del intervalo lineal del método. Pruebas similares aplicadas al cuarto punto indican, al compararlos, que probablemente deberían formar parte del intervalo lineal, y un analista podría usar métodos de regresión lineal en el intervalo de absorbancia 0 - 0.6 con confianza.

- Este ejemplo también recalca la importancia del trazado de una curva de calibración en vez de realizar un simple cálculo de regresión lineal. El último cálculo produce un coeficiente de correlación de 0.9952, aunque la inspección de la gráfica muestre obviamente que es curva. La regresión lineal muestra que los residuos y tienen signos (en orden creciente de  $x$ ) de - - - + + + + - -; la suma de los cuadrados de los residuos es 9.50. Es claramente más improbable que el orden de los signos sea aleatorio. Cuando la curva  $y = 4x - 0.1x^2$  se ajusta a los datos, el orden de los signos de los residuos es: + - - + + - 0 - + - +, y la suma de los cuadrados es sólo 1.29. Existe, entonces, evidencia de que los puntos se ajustan mejor por la curva que por la línea recta. Para los valores de  $y$ , 5, 16 y 27 la línea recta predice concentraciones de 1.15, 4.83 y 8.51 unidades respectivamente, mientras que la curva predice concentraciones de 1.29, 4.51 y 8.60 unidades respectivamente. A la luz de los análisis de los residuos, los valores últimos son ciertamente más fiables. No debe pasarse por alto, sin embargo, que otras curvas podrían ajustar los puntos experimentales aun mejor que la que se sugiere aquí: el análisis con computador muestra que "el mejor" ajuste cuadrático es  $y = 0.86 + 3.97x - 0.098x^2$ .

El ejemplo también muestra que, en los casos donde existe un razonable número de puntos de calibración, y donde los errores aleatorios no son muy grandes, el tratamiento de la curva como una serie de segmentos rectos cortos, es un método aproximado atractivo. Si se dibujan diez segmentos lineales, es decir, se utilizan todos los puntos del calibrado, las concentraciones calculadas por interpolación lineal para señales de 5, 16 y 27 unidades son 1.36, 4.50 y 8.65 unidades respectivamente. Incluso si sólo se dibujan cinco segmentos, es decir, sólo se usan los puntos  $x = 0, 2, 4$ , etc., las concentraciones predichas son 1.32, 4.37 y 8.51 unidades. En vista de la facilidad de cálculo y los derivados de la curva ajustada, los atractivos de este método son evidentes.

## CAPÍTULO 6

- La media es 9.96 ml, y la mediana (el promedio de las mediciones que ocupan la segunda y tercera posiciones) es 9.90 ml. La prueba  $Q$  muestra que el valor 10.20 no se puede rechazar a  $P = 0.05$ . Si *fuera* rechazada, la media cambiaría a 9.88 ml y la mediana a 9.89 ml; la mediana es insensible a la presencia de valores anómalos.
- (a) Se usa la prueba de los signos: comparados con la mediana proporcionan, en los valores experimentales, los signos - + 0 + - + + +. En efecto, de este modo tenemos 8 lecturas, 6 con signo +. La probabilidad de 6 signos en 8 lecturas es 0.29, mucho mayor que 0.05. Por lo tanto se mantiene la hipótesis nula de procedencia de los datos de una población simétrica con contenido mediano de azufre de 0.10%.

(b) En la prueba de rangos y signos, se desprecia de nuevo el valor cero. Los datos restantes comparados con la mediana proporcionan las siguientes diferencias ordenadas: -0.01, 0.01, 0.01, -0.02, 0.02, 0.02, 0.04, 0.07. Las posiciones con sus signos, teniendo en cuenta los empates, son por tanto -2, 2, 2, -5, 5, 5, 7, 8. El total de posiciones negativas (-) es 7 pero, a  $P = 0.05$ , el valor crítico para  $n = 8$  es solamente de 3. Por lo tanto, se mantiene de nuevo la hipótesis nula.

(c) Para utilizar una prueba de rangos, calculamos la media ( $\bar{x}$ ), que es 0.1156 y el recorrido ( $w$ ), 0.09. Entonces  $T_l$  está dado por  $|\bar{x} - V|/w = 0.173$ . El valor crítico en este caso es 0.255, de modo que por tercera vez se mantiene la hipótesis nula. En contra de las apariencias observadas a simple vista no se puede decir, quizás, que la muestra de combustible no cumpla la especificación.

- Este problema se puede resolver también utilizando la prueba de signos, y el de rangos y signos. Si se restan los valores e.i.d. de los valores i.d.r., los signos de las diferencias son: + - + + + + + 0 +. En efecto, 9 resultados, con 8 signos + y un signo -. La probabilidad de que haya 8 signos + en 9 resultados es solamente de 0.04, por tanto al nivel de significación  $P = 0.05$  se puede rechazar la hipótesis nula: los resultados *son* significativamente diferentes. En la prueba de rangos y signos, el valor asociado a las posiciones negativas es sólo de 2.5. Este número es muy inferior al valor crítico de 5, por tanto se rechaza otra vez la hipótesis nula, y se confirma la diferencia significativa entre los dos conjuntos de resultados. Observe, no obstante, que los resultados de la prueba de rangos y signos deberían interpretarse con precaución si hay muchas posiciones empatadas, como ocurre en este ejercicio y en el anterior.
- El nuevo método proporciona una media de 64.7 mg/100 ml y un intervalo de 2.1 mg/100 ml: las cifras correspondientes para el método estándar son 65.86 y 0.9 mg/100 ml. De estas cifras, se ve claramente que  $T_a$  es 0.773, mayor que el valor crítico de 0.61, de manera que las medias *difieren* significativamente. La razón de posiciones  $F_R$  es 2.33, comparada con el valor crítico de 3.2, de manera que las precisiones de los métodos no son significativamente diferentes. (La prueba  $F$  da el mismo resultado.)
- Si los valores se disponen en orden ascendente, se obtiene que la mediana es 23.5. Los valores individuales, para el uso de las barras de grafito, tienen por lo tanto los signos + + + - - - - + +. De este modo existen tres rachas. Pero para  $N = M = 5$  el valor crítico es 3, con lo cual se podría decir que la sucesión dada se ha obtenido al azar.
- Este problema se resuelve fácilmente mediante la prueba  $U$ . Si se espera que los valores de "cerveza estándar" sean mayores que los valores de "cerveza rubia", debería de ser pequeño el número de valores de cerveza rubia que son mayores que los de cerveza estándar. Es realmente 4.5, contando 0.5 para el empate que aparece. El valor crítico en una prueba de *una cola* es 5, con lo cual podemos rechazar la hipótesis nula al nivel de  $P = 0.05$  y afirmar que una cerveza rubia en particular produce niveles de alcohol en sangre inferiores a los de una cerveza estándar en concreto! Sin embargo, el resultado en la prueba de ejecución rápida de Tukey,  $T = 5.5$ , no llega a alcanzar el valor crítico de 6 (una cola,  $P = 0.05$ ). Por tanto, se deben tratar los resultados con precaución. Se necesitan más datos. La prueba de Siegel-Tukey para la variabilidad



proporciona un estadístico de la prueba de 10, muy por encima del valor crítico de 2. Por tanto, la hipótesis nula de variabilidades similares se debe aceptar.

7. Las posiciones se muestran a continuación:

Instrumento	A	B	C	D	E	F	G
Ordenación del estudiante	3	1	5	4	7	6	2
Ordenación del personal académico	5	3	6	2	4	7	1
$d$ :	-2	-2	-1	2	3	-1	1
$d^2$ :	4	4	1	4	9	1	1

Así pues, es  $1 - (6 \times 24)/(7 \times 48) = 0.571$ . Éste se encuentra por debajo de valor crítico 0.786 ( $P = 0.05$  y  $n = 7$ ), con lo cual no parece que los estudiantes y el personal académico están de acuerdo —la correlación no es significativa.

8. Si los valores  $x$  son las distancias, y los valores  $y$  los niveles de mercurio, entonces el método de Theil da  $a = 2.575$ , y  $b = 0.125$ . Estos valores están sorprendentemente próximos a los niveles (paramétricos) de “mínimos cuadrados” de 2.573 y -0.122.

9. Para probar  $\bar{x} = 1.0$  y  $\sigma = 0.2$ , escribimos  $z = (x - 1.0)/0.2$ , y obtenemos 1.5, 2.5, -1.5, -0.5, 0, 0.5, -1.0, 4.0, -3.0 y 1.5. Si dibujamos dichos valores frente a la función de distribución normal estándar, resulta que dan lugar a un trazo que casi es simétrico, pero mucho más ancho que la curva estándar, indicando que la media supuesta no se aleja demasiado de la verdadera, pero la desviación estándar estimada es demasiado pequeña. La máxima diferencia es +0.335 en  $z = 1.5$ . Este valor es mucho más alto que el crítico de 0.262, lo que confirma que se debe rechazar la hipótesis nula, y que los valores no se ajustan a esta distribución normal concreta. Para probar otra distribución normal, calculamos las estimaciones de la media y la desviación estándar que son respectivamente 1.08 y 0.41. Esto proporciona los valores  $z$  de 0.54, 1.02, -0.93, -0.44, -0.20, 0.05, -0.68, 1.76, -1.66 y 0.54. La máxima diferencia, en este caso, es sólo de ca. 0.11 en  $z = 0.54$ , por tanto se acepta la hipótesis nula; los datos se ajustan muy bien a esta distribución normal.

10. Este problema se resuelve sustituyendo las concentraciones de níquel reales con valores de posición, con los resultados siguientes:

Muestra	Valores de posición					
1	1	8	5	15	4	6
2	2	17	10.5	13	3	7
3	14	18	9	10.5	12	16

Observe que se da un empate y, por tanto, se tratará de la manera habitual. Las sumas de los tres conjuntos de posiciones son:  $R_1 = 39$ ,  $R_2 = 52.5$ ,  $R_3 = 79.5$ . Estos tres números suman 171, igual a  $\frac{1}{2}N(N + 1)/2$ , donde  $N = 18$ , es el número total de mediciones. El valor del estadístico chi-cuadrada, está dado por:

$$\chi^2 = \frac{12}{342 + 18} \left( \frac{1521}{6} + \frac{2756.25}{6} + \frac{6320.25}{6} \right) - 3(18 + 1) = 4.97$$

Este valor es menor que el crítico para  $\chi^2$  de 5.99 ( $P = 0.05$ , 2 grados de libertad —tabla A.5). Por tanto, debemos mantener la hipótesis nula, y concluir que no hay diferencia significativa entre los niveles de níquel en las tres muestras de aceite crudo.

## CAPÍTULO 7

1. Un ejemplo de ANOVA de dos factores con replicación.

Fuente de variación	Suma de cuadrados	g.l.	Cuadrado medio
Entre filas	5.06778	2	2.3
Entre columnas	0.18778	2	0.0939
Interacción	0.10222	4	0.0256
Residual	0.36500	9	0.0406
Total	5.72278	17	

El cuadrado medio residual es mayor que el cuadrado medio de la interacción, por eso no existe interacción significativa muestra-laboratorio. Para probar si hay diferencia significativa entre los laboratorios se compara el cuadrado medio entre columnas con el cuadrado medio residual, dando:

$$F_{2,9} = 0.0939/0.0406 = 2.31$$

El valor crítico de  $F_{2,9}$  ( $P = 0.05$ ) es 4.2556, con lo cual el valor calculado no es significativo.

2. El vértice 1 debería ser rechazado al formar el nuevo simplex.

	Factores					Respuesta
	A	B	C	D	E	
Vértice 2	6.0	4.3	9.5	6.9	6.0	8
Vértice 3	2.5	11.5	9.5	6.9	6.0	10
Vértice 5	2.5	4.3	9.5	9.7	6.0	11
Vértice 6	2.5	4.3	9.5	6.9	9.6	9
Vértice 7	3.3	6.7	12.5	7.7	7.0	12
Suma	16.8	31.1	50.5	38.1	34.6	
Suma/ $n$	3.4	6.2	10.1	7.6	6.9	
Vértice rechazado (es decir, 1)	1.0	3.0	2.0	6.0	5.0	
Desplazamiento	2.4	3.2	8.1	1.6	1.9	
Vértice 8	5.8	9.4	18.1	9.2	8.8	

(Todos los valores de la tabla anterior se expresan con una cifra decimal.)

3. Un ejemplo de ANOVA de dos factores sin replicación. Primero se resta 16 de cada valor.

Fuente de variación	Suma de cuadrados	g.l.	Cuadrado medio
Entre filas	0.011092	3	0.00370
Entre columnas	0.012014	2	0.00601
Residual	0.028186	6	0.00470
Total	0.051292	11	

(a) El cuadrado medio entre soluciones (es decir, filas) es más pequeño que el cuadrado medio residual, de manera que no es significativo.

(b) Comparando los cuadrados medios entre métodos (es decir, columnas) y residuales proporciona  $F_{2,6} = 1.28$ , que no es significativo ( $P = 0.05$ ).

4. Un ejemplo de ANOVA de dos factores con replicación.

Fuente de variación	Suma de cuadrados	g.l.	Cuadrado medio
Entre filas	20.8768749	3	6.959
Entre columnas	0.14062595	1	0.141
Interacción	0.11187458	3	0.037
Residual	9.92499971	8	1.241
Total	31.0543752	15	

No existe interacción significativa o diferencia entre los métodos ya que en ambos casos el correspondiente cuadrado medio es menor que el cuadrado medio residual.

5. Utilizando la siguiente notación:

G = gas de purga (Ar—bajo, Ar + O<sub>2</sub>—alto)

T = tiempo de purga (20 segundos—bajo, 60 segundos—alto)

Q = temperatura de la célula de sílice (900C—bajo, 1000°C—alto)

los efectos calculados son:

Un factor	Efecto
G	18.25
T	-12.25
Q	55.25
Dos factores	
GT	13.75
TQ	-5.25
CQ	-0.75
Tres factores	
GTQ	3.75

Una factor	Efecto
A	-0.0215
C	0.0005
T	-0.0265
Dos factores	
AC	-0.0005
CT	0.0025
AT	-0.0065
Tres factores	
ACT	-0.0005

7. Utilizando la siguiente notación:

N = concentración de ácido nítrico (0.1M – bajo, 1.0 M – alto)

L = concentración de lantano (0M – bajo, 0.1M – alto)

los efectos calculados son:

Un factor	Efecto	Suma de cuadrados	g.l
N	9.5	180.5	1
L	22.5	1012.5	1
Interacción			
NL	3.5	25.5	1

Para probar si es significativa la interacción se compara el cuadrado medio de la interacción con el cuadrado medio de los residuos, dando  $F_{1,4} = 24.5/4 = 6.125$ . El valor crítico de  $F_{1,4}$  es 7.709 ( $P = 0.05$ ), de manera que no existe interacción significativa.

# Apéndice 1 Resumen de pruebas estadísticas

Las siguientes tablas resumen las pruebas estadísticas descritas en este libro. Las pruebas se agrupan de acuerdo con sus aplicaciones, con comentarios y breves detalles en cada caso, y una referencia a la página en la que pueden encontrarse otros detalles. Es de esperar que estas tablas ayuden al lector en la elección de la prueba estadística adecuada para un problema concreto.

## (1) Pruebas sobre promedios

Nombre de la prueba o estadístico de la prueba	Utilizado para probar si:	Referencia en página	Comentarios
$t$	(a) una media muestral difiere de un valor estándar (b) las medias de dos muestras difieren (c) las muestras por parejas difieren	53 55 58	Supone que las poblaciones son normales
Prueba de signos	(a) una mediana muestral difiere de un valor estándar (b) las muestras por parejas difieren	138 139	No paramétrico
$T_l$	(a) una media muestral difiere de un valor estándar	142	Supone que la población es normal; más rápido que la prueba $t$
$T_d$	(a) las medias de dos muestras difieren	143	Supone que las poblaciones son normales; más rápido que la prueba $t$
$L$ (Prueba de recorridos de Lord)	(a) las medias de dos muestras difieren	143	Supone que las poblaciones son normales; más rápido que la prueba $t$
Prueba de rangos y signos de Wilcoxon	(a) una mediana muestral difiere de un valor estándar (b) la muestras por parejas difieren	144 145	No paramétrica, pero supone que la población es simétrica; requiere $n \geq 5$ para pruebas de una cola y $n \geq 6$ para dos colas
Prueba de suma de rangos de Wilcoxon	(a) las medianas de dos muestras difieren	146	No paramétrica; tamaños de muestra mínimos: prueba de una cola $n_1 = n_2 = 3$ ; prueba de dos colas, $n_1 = 3, n_2 = 5$
Prueba $U$ de Mann-Whitney	(a) las medianas de dos muestras difieren	148	Versión más simple de la prueba de suma de rangos de Wilcoxon
Prueba rápida de Tukey	(a) las medianas de dos muestras difieren	148	Alternativa muy simple a las pruebas de suma de rangos y $U$ —no necesita tablas estadísticas
Prueba de Kruskal-Wallis	(a) las medianas de tres o más muestras difieren	150	No paramétrica; utiliza el método $\chi^2$ en forma modificada
Prueba de Friedman	(a) las muestras emparejadas difieren	151	No paramétrica; método $\chi^2$ en forma modificada

## (2) Pruebas sobre variabilidad

Nombre de la prueba o estadístico de la prueba	Utilizado para probar si:	Referencia en página	Comentarios
$F$	las varianzas de dos muestras difieren	60	Supone normalidad
$F_R$	los recorridos de dos muestras difieren	143	Más sencillo de realizar que la prueba $F$ . Supone normalidad
Prueba de Siegel-Tukey	las variabilidades de dos muestras difieren	149	No paramétrica; modificada de una prueba de suma de rangos de Wilcoxon
(3) Pruebas de racha			
Wald-Wolfowitz	una secuencia es aleatoria	141	Para números mínimos de datos véase la tabla A.8
(4) Pruebas sobre correlación			
Coefficiente de correlación producto-momento, $r$	los datos están correlacionados linealmente	104	Supone que los errores se distribuyen normalmente
Coefficiente de correlación ordinal de Spearman, $\rho$	los datos están correlacionados linealmente	153	Más simple de calcular que $r$ , no paramétrica
(5) Pruebas de bondad del ajuste			
$\chi^2$ (chi-cuadrada)	las frecuencias observadas difieren de las frecuencias esperadas	71	Requiere un tamaño de muestra $\geq 50$ y frecuencias individuales (observadas) $\geq 5$
Kolmogorov-Smirnov	los datos se adaptan a una distribución especificada	156	Puede utilizarse para muestras pequeñas; especialmente útil al probar normalidad

## Apéndice 2 Tablas estadísticas

Se presentan para la conveniencia del lector las siguientes tablas, y para su uso en este libro en pruebas estadísticas simples, ejemplos y ejercicios. Están presentadas en un formato compatible con las necesidades de químicos analíticos: el nivel de significación  $P = 0.05$  ha sido utilizado en muchos casos, y se ha supuesto que el número de mediciones disponible es suficientemente pequeño. Excepto donde no se establezca lo contrario, estas tablas abreviadas han sido obtenidas, con autorización de *Elementary Statistics Tables* de Henry R. Neave, publicadas por George Allen & Unwin Ltd. (Tablas A.1-A.3, A.5-A.7 y A.12-A.16). Se remite a estas fuentes a los lectores que requieran números de mediciones no recogidos en las tablas.

Tabla A.1 La distribución  $t$

Valor de $t$ para un intervalo de confianza de Valor crítico de $ t $ para valores de $P$ de Número de grados de libertad	90%	95%	98%	99%
1	6.31	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
12	1.78	2.18	2.68	3.05
14	1.76	2.14	2.62	2.98
16	1.75	2.12	2.58	2.92
18	1.73	2.10	2.55	2.88
20	1.72	2.09	2.53	2.85
30	1.70	2.04	2.46	2.75
50	1.68	2.01	2.40	2.68
$\infty$	1.64	1.96	2.33	2.58

Los valores críticos de  $|t|$  son adecuados para una prueba de *dos* colas. Para una prueba de *una* cola el valor se toma de la columna para dos veces el valor de  $P$  deseado, por ejemplo, para una prueba de una cola,  $P = 0.005$ , 5 grados de libertad, el valor crítico se lee de la columna  $P = 0.10$  y es igual a 2.02.

Tabla A.2 Valores críticos de $F$ para una prueba de una cola ( $P = 0.05$ )													
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	12	15	20
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.678	4.619	4.558
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347	3.284	3.218	3.150
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.936
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.913	2.845	2.774
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.475	2.403	2.328
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.155
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124

$\nu_1$  = número de grados de libertad del numerados y  $\nu_2$  = número de grados de libertad del denominador.

Tabla A.3 Valores críticos de $F$ para una prueba de dos colas ( $P = 0.05$ )													
$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	12	15	20
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17
4	12.22	10.65	9.979	9.605	9.364	9.197	9.074	8.980	8.905	8.844	8.751	8.657	8.560
5	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428	6.329
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.269	5.168
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568	4.467
8	7.571	6.059	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101	3.999
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769	3.667
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522	3.419
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.430	3.330	3.226
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.177	3.073
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.153	3.053	2.948
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.050	2.949	2.844
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862	2.756
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.788	2.681
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.825	2.723	2.616
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.769	2.667	2.559
19	5.922	4.508	3.903	3.559	3.333	3.172	3.051	2.956	2.880	2.817	2.720	2.617	2.509
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.837	2.774	2.676	2.573	2.464

$\nu_1$  = número de grados de libertad del numerador y  $\nu_2$  = número de grados de libertad del denominador.

Tabla A.4 Valores críticos de  $Q$  ( $P = 0.05$ )

Tamaño de muestra	Valor crítico
4	0.831
5	0.717
6	0.621
7	0.570
8	0.524
9	0.492
10	0.464

Tomados de King, E. P., *J. Am. Statist. Assoc.*, 1958, **48**, 531, con autorización de American Statistical Association.

Tabla A.5 Valores críticos de  $\chi^2$  ( $P = 0.05$ )

Número de grados de libertad	Valor crítico
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
6	12.59
7	14.07
8	15.51
9	16.92
10	18.31

Tabla A.6 Números aleatorios

02484	88139	31788	35873	63259	99886	20644	41853	41915	02944
83680	56131	12238	68291	95093	07362	74354	13071	77901	63058
37336	63266	18632	79781	09184	83909	77232	57571	25413	82680
04060	46030	23751	61880	40119	88098	75956	85250	05015	99184
62040	01812	46847	79352	42478	71784	65864	84904	48901	17115
96417	63336	88491	73259	21086	51932	32304	45021	61697	73953
42293	29755	24119	62125	33717	20284	55606	33308	51007	68272
31378	35714	00941	53042	99174	30596	67769	59343	53193	19203
27098	38959	49721	69341	40475	55998	87510	55523	15549	32402
66527	73898	66912	76300	52782	29356	35332	52387	29194	21591
61621	52967	40644	91293	80576	67485	88715	45293	59454	76218
18798	99633	32948	49802	40261	35555	76229	00486	64236	74782
36864	66460	87303	13788	04806	31140	75253	79692	47618	20024
10346	28822	51891	04097	98009	58042	67833	23539	37668	16324
20582	49576	91822	63807	99450	18240	70002	75386	26035	21459
12023	82328	54810	64766	58954	76201	78456	98467	34166	84186
48255	20815	51322	04936	33413	43128	21643	90674	98858	26060
92956	09401	58892	59686	10899	89780	57080	82799	70178	40399
87300	04729	57966	95672	49036	24993	69827	67637	09472	63356
69101	21192	00256	81645	48500	73237	95420	98974	36036	21781
22084	03117	96937	86176	80102	48211	61149	71246	19993	79708
28000	44301	40028	88132	07083	50818	09104	92449	27860	90196
41662	20930	32856	91566	64917	18709	79884	44742	18010	11599
91398	16841	51399	82654	00857	21068	94121	39197	27752	67308
46560	00597	84561	42334	06695	26306	16832	63140	13762	15598

**Tabla A.7** La prueba de signos

La tabla utiliza la distribución binominal con  $P = 0.5$  para proporcionar las probabilidades de  $r$  éxitos o menos para  $n = 4 - 15$ . Estos valores corresponden a una prueba de signos de una cola y se deben duplicar para una prueba de dos colas.

$n$	$r = 0$	1	2	3	4	5	6	7
4	0.063	0.313	0.688					
5	0.031	0.188	0.500					
6	0.016	0.109	0.344	0.656				
7	0.008	0.063	0.227	0.500				
8	0.004	0.035	0.144	0.363	0.637			
9	0.002	0.027	0.090	0.254	0.500			
10	0.001	0.011	0.055	0.172	0.377	0.623		
11	0.001	0.006	0.033	0.113	0.274	0.500		
12	0.000	0.003	0.019	0.073	0.194	0.387	0.613	
13	0.000	0.002	0.011	0.046	0.133	0.290	0.500	
14	0.000	0.001	0.006	0.029	0.090	0.212	0.395	0.605
15	0.000	0.000	0.004	0.018	0.059	0.151	0.304	0.500

**Tabla A.8** La prueba de rachas de Wald-Wolfowitz

$N$	$M$	A $P = 0.05$ , el número de rachas es significativo si es:	
		menor que	mayor que
2	12-20	3	NA
3	6-14	3	NA
3	15-20	4	NA
4	5-6	3	8
4	7	3	NA
4	8-15	4	NA
4	16-20	5	NA
5	5	3	9
5	6	4	9
5	7-8	4	10
5	9-12	4	NA
5	13-18	5	NA
6	6	4	10
6	7-8	4	11
6	9-12	5	12
6	13-18	6	NA
7	7	4	12
7	8	5	12
7	9	5	13
7	10-12	6	13
8	8	5	13
8	9	6	13
8	10-11	6	14
8	12-15	7	15

Adaptada de Swed, F. S. y Eisenhar, C., *Ann. Math. Statist.* 1943, **14**, 66, con autorización del Institute of Mathematical Statistics. La prueba no puede aplicarse a datos con  $N, M$  menores que los números dados, o a casos marcados con NA.

**Tabla A.9** La prueba  $T_l$

Valores críticos para pruebas de una y dos colas a  $P = 0.05$

$n$	Prueba de una cola	Prueba de dos colas
2	3.175	6.353
3	0.885	1.304
4	0.529	0.717
5	0.388	0.507
6	0.312	0.399
7	0.263	0.333
8	0.230	0.288
9	0.205	0.255
10	0.186	0.230

**Tabla A.10** La prueba  $T_d$  y la prueba de recorridos de Lord

Valores críticos de  $T_d$  y  $L$  a  $P = 0.05$

$n_1 = n_2$	$T_d$	$L$
2	3.43	1.71
3	1.27	0.64
4	0.81	0.41
5	0.61	0.31
6	0.50	0.25
7	0.43	0.21
8	0.37	0.19
9	0.33	0.17
10	0.30	0.15

Las tablas A.9 y A.10 están tomadas de Lord, E., *Biometrika*, 1947, **34**, 66, con autorización de Biometrika Trustees.

**Tabla A.11** La prueba sustituta de la prueba  $F$  ( $F_R$ )

Valores críticos de  $F_R$  para pruebas de una y dos colas a  $P = 0.05$

Número de mediciones en numerador y denominador	Prueba de una cola	Prueba de dos colas
2	12.7	25.5
3	4.4	6.3
4	3.1	4.0
5	2.6	3.2
6	2.3	2.8
7	2.1	2.5
8	2.0	2.3
9	1.9	2.2
10	1.9	2.1

Adaptada de Link, R. F., *Ann. Math. Statist.*, 1950, **21**, 122 con autorización del Isititute of Mathematical Statistics.

**Tabla A.12** La prueba de rangos y signos de Wilcoxon

Valores criticos para el estadístico a  $P = 0.05$

$n$	Prueba de una cola	Prueba de dos colas
5	0	NA
6	2	0
7	3	2
8	5	3
9	8	5
10	10	8
11	13	10
12	17	13
13	21	17
14	25	21
15	30	25

La hipótesis nula puede rechazarse cuando el estadístico de la prueba sea  $\leq$  que el valor tabulado. NA indica que la prueba no puede ser plicada.

**Tabla 14** El coeficiente de correlación ordinal de Spearman

Valores críticos de a  $P = 0.05$

$n$	Prueba de una cola	Prueba de dos colas
5	0.900	1.000
6	0.829	0.886
7	0.714	0.786
8	0.643	0.738
9	0.600	0.700
10	0.564	0.649
11	0.536	0.618
12	0.504	0.587
13	0.483	0.560
14	0.464	0.538
15	0.446	0.521
16	0.429	0.503
17	0.414	0.488
8	0.401	0.472
19	0.391	0.460
20	0.380	0.447

**Tabla A.13** La prueba de suma de rangos de Wilcoxon; la prueba  $U$  de Mann-Withney

Valores criticos de  $U$  o el menor entre  $T_1$  y  $T_2$  a  $P = 0.05$

$n_1$	$n_2$	Prueba de una cola	Prueba de dos colas
3	3	0	NA
3	4	0	NA
3	5	1	0
3	6	2	1
4	4	1	0
4	5	2	1
4	6	3	2
4	7	4	3
5	5	4	2
5	6	5	3
5	7	6	5
6	6	7	5
6	7	8	6
7	7	11	8

La hipótesis nula puede rechazarse cuando  $U$  o el valor de  $T$  más pequeño sea  $\leq$  que el valor tabulado. NA indica que la prueba no puede ser aplicada.

**Tabla A.15** La prueba de bondad del ajuste de Kolmogorov

Los valores críticos para la prueba de una y dos colas a  $P = 0.05$ . Los valores adecuados se comparan con la diferencia máxima entre las curvas de frecuencia acumulada experimental y teórica, tal y como se describe en el texto.

$n$	Prueba de una cola	Prueba de dos colas
1	0.950	0.975
2	0.776	0.842
3	0.636	0.708
4	0.565	0.624
5	0.509	0.563
6	0.468	0.519
7	0.436	0.483
8	0.410	0.454
9	0.388	0.430
10	0.369	0.409
11	0.352	0.392
12	0.338	0.375
13	0.326	0.361
14	0.314	0.349
15	0.304	0.338
16	0.295	0.327
17	0.286	0.318
8	0.278	0.309
19	0.271	0.301
20	0.265	0.294

**Tabla A.16** La prueba de Kolmogorov para probar la normalidad

Los valores críticos para pruebas de una y dos colas a  $P = 0.05$ . Los valores adecuados se comparan con la diferencia máxima entre las curvas de frecuencia acumulada experimental y teórica, tal y como se describe en el texto.

<i>n</i>	Prueba de una cola	Prueba de dos colas
3	0.367	0.376
4	0.345	0.375
5	0.319	0.343
6	0.297	0.323
7	0.280	0.304
8	0.265	0.288
9	0.252	0.274
10	0.241	0.262
11	0.231	0.251
12	0.222	0.242
13	0.215	0.234
14	0.208	0.226
15	0.201	0.219
16	0.195	0.213
17	0.190	0.207
8	0.185	0.202
19	0.181	0.197
20	0.176	0.192

# Índice de materias

absorbancia, 12, 32, 35	análisis de tamaños de partículas, 2
ajuste de curvas, 115, 118, 119, 126, 141	análisis enzimático, 14, 15, 165, 166
albúmina en suero, determinación de, 1, 13, 25	análisis gravimétrico, 87
aleatoriedad, prueba para la, 142, 143	análisis potenciométrico, 35, 36
aleatorización, 148, 149, 160	análisis volumétrico, 7-10, 33, 35, 87, 122
American Society for Testing and Materials (ASTM), 7, 17	analizadores centrifugos 170-171
análisis automático, 88, 90, 169, 170	aprendizaje no supervisado, métodos de reconocimiento de pautas, 172-173, 176-180
análisis clínicos, 188	aprendizaje supervisado, métodos de reconocimiento de pautas, 172, 175, 176
análisis cualitativo, 1	ARTHUR, método de reconocimiento de pautas, 179
análisis cuantitativo, 1-3	astigmatismo, 11
análisis de componentes principales, 172, 174, 179	
análisis de la varianza (ANOVA), 52-56, 67-70, 75, 76, 84, 137, 147-149, 179	barras de error, 111
aritmética de cálculos, 56, 57	blanco, 10, 89, 90, 101-102, 108
cuadrado medio residual en, 150-152, 157	bloque(s), 148, 149, 152
cuadrados medios en, 53-56, 150-152, 157, 162, 163	bondad del ajuste, 141-145, 198, 205
de un factor, 52-56, 75, 150, 153	British Standards Institution (BSI), 7, 17
diferencias significativas en, 55-56	bureta, 7-9
de dos factores, 149-158, 162	búsqueda de variable alternativa, 164
mínima diferencia significativa en, 82, 83	
para comparar varias medias, 52-56	calculadoras, 15, 21, 92, 96, 99, 139
suma de cuadrados en, 55, 56, 150, 151, 157, 158, 162, 163	capacidad de procesos, 79, 80, 84
variación dentro de muestras en, 53-56, 69, 70, 76, 79	centro de gravedad ponderado, 111-114
variación entre bloques en, 149, 151	centro de gravedad, de puntos en gráficas de calibración, 90, 95, 100, 106, 111-113
variación entre muestras en, 53-56, 69, 70, 75, 76, 78	cifras significativas, 23, 31, 92, 98, 151
variación entre tratamientos en, 149, 153	clasificación binaria, 172-173
variación total en, 55, 56	clasificador, 172-173
	censura de valores en ensayos de colaboración, 76





método de la máquina de aprendizaje, 175  
método de mínimos cuadrados, 16, 95, 118, 140  
método de Theil para rectas de regresión, 139-141  
método del *K* entorno más próximo (KNN), 175-176  
métodos de distribución libre, *véase* métodos no paramétricos  
métodos de interpolación en regresión no lineal, 119  
métodos de Kolmogorov-Smirnov, 60, 141-145, 198, 205  
métodos de la transformada de Fourier, 88  
métodos de ordenación, 129-141  
métodos de regresión no ponderados, 100, 102, 105, 110-113, 116, 118  
métodos de regresión ponderada, 110-114, 117, 118  
métodos de regresión, supuestos utilizados en, 90, 110, 111  
    lineal, 15, 16, 46, 51, 87-119, 125, 126  
    no lineal, 15, 17, 88, 93, 113, 119  
    no paramétrica, 139-141  
    para comparar métodos analíticos, 106-110  
métodos electroquímicos de análisis, 87, 103, 114  
métodos modificados en la optimización simplex, 169  
métodos no paramétricos, 50, 96, 122-127, 129-144, 177, 197, 198  
métodos radioquímicos de análisis, 87  
métodos robustos, 141  
métodos térmicos de análisis, 87  
microcomputadores, 11, 15, 17, 88, 92, 99, 108, 111, 113, 118, 123, 139, 141, 170-171  
microprocesadores, *véase* microcomputadores  
minicomputadores, 16  
mínima diferencia significativa, en ANOVA, 54-55  
monocromadores, errores sistemáticos debidos a, 11-13  
muestra aleatoria, 68  
muestra bruta, 68  
muestra de mediciones, 23  
muestreo a granel, 68  
muestreo, 65, 81-84, 88

naranja de metilo, error de indicador debido a, 10  
National Bureau of Standards (NBS), 13, 17, 18  
National Physical Laboratory (NPL), 13  
niveles de factores experimentales, 73, 74, 148, 153, 154, 159-161, 165-169, 169, 170  
niveles de significación, 40, 41  
nudos en funciones a trozos (*spline*), 133

optimización simplex, 17, 168-170, 208, 212

optimización, 14, 17, 88, 147, 158-172  
ordenación alterna por pares, 134, 135  
ordenada en el origen, de la gráfica de calibración lineal, 89, 96, 98, 99, 102-103, 107-110, 110-111, 112, 139-141

pendiente de la gráfica de calibración lineal, 89, 96, 98, 99, 102-103, 107-110, 111, 112, 125, 139-141

periodicidad, efectos en el muestreo, 96  
    de signos + y -, 127

pesada, 7, 8  
    efectos atmosféricos en la, 8, 9  
    por diferencia, 7, 9, 37  
    recipiente de, 7

pipeta, 7-9  
plano de decisión, 173  
plantilla en V, 82-84  
población de mediciones, 23, 67  
ponderaciones, de puntos de regresión ponderada, 111, 112, 113, 117, 118  
posiciones empatadas, 131, 133-136, 139  
potencia de una prueba estadística, 62, 63, 125, 129, 145  
precisión dentro de rachas, 4, 5  
precisión, 4-6, 26, 50, 57, 82, 159  
prejuicios en la experimentación, 11  
presentación de resultados, 31, 32  
problemas de especiación, 108  
procesadores de texto, 43  
propagación de errores aleatorios, 14, 22, 33-36  
propagación de errores sistemáticos, 37  
prueba bilateral, *véase* prueba de dos colas  
prueba chi-cuadrada, 85, 86, 136, 137, 141-143, 198, 201S  
    determinación de,  
prueba de dos colas, 46, 47, 49, 63, 94, 124, 125, 128, 130, 131, 134, 135, 143, 197, 199-200  
prueba de entereza, 73  
prueba de Friedman, 136, 137, 197  
prueba de Kruskal-Wallis, 135, 136, 197  
prueba de rachas de Wald-Wolfowitz, 126, 127, 198, 202  
prueba de rangos y signos de Wilcoxon, 129, 131, 135, 204  
prueba de rangos y signos, 129-131, 135  
prueba de recorridos de Lord, 128, 197  
prueba de Siegel-Tukey, 134, 135, 198  
prueba de signos, 123-126, 129, 197, 202  
prueba de suma de rangos de Wilcoxon, 131-134, 145, 197, 204  
prueba de suma de rangos, 131-134, 145  
prueba de una cola, 46-49, 55, 59, 63, 128, 130, 132, 134, 197, 199, 209-206  
prueba *F* sustituto, 128, 129, 203

Prueba *F*, 32, 47-49, 63, 70, 76-78, 128, 129, 134, 151, 152, 163, 200  
prueba *Q* para valores anómalos, 50, 51, 78, 201  
prueba *t* por parejas, 45, 46, 124, 130  
prueba rápida de Tukey, 133, 134, 197  
prueba *t*, 32, 41-46, 49, 58, 63, 67, 94, 116, 124, 128, 197, 203  
prueba *U* de Mann-Whitney, 133, 134, 197, 204  
pruebas de significación, 32, 40-66, 78  
    errores en el uso de, 61-63  
    problemas en el uso secuencial, 63  
papel de probabilidad normal, 58, 59, 61, 142

*Q* de Dixon, 50, 51, 78, 201  
quimiometría, 15, 170

rachas de signos + y -, 115, 126, 127, 198, 202  
reconocimiento de pautas, 17, 147, 170-179  
recorrido intercuartílico, 123, 127  
recorrido, 79, 80, 84, 127-129, 198  
recta de regresión de *x* sobre *y*, 95  
recta de regresión de *y* sobre *x*, 95, 110  
recta de regresión, 101-102  
redondeo de resultados, 31, 32, 98, 115  
regresión curvilínea, *véase* regresión no lineal  
regresión no lineal, 15, 17  
repetitividad, 4-6  
réplicas, en diseño de experimentos, 155  
reproducibilidad, 4-8  
residuos (de) *y*, en gráficas de calibración, 95-97, 115, 116, 126  
residuos, en cálculos de regresión, *véase* residuos y superficies de respuesta en optimización, 164, 165, 167

semejanza en el reconocimiento de pautas, 211  
sensibilidad, 103-104  
señal de fondo, *véase* blanco,  
sesgo, *véase* error sistemático,  
SIMCA, método de reconocimiento de pautas, 176, 178  
sistemas expertos, 16  
*spline*, *véase* funciones a trozos

*splines* (funciones a trozos) cúbicos, 118  
sucesiones, de signos + y -, 115  
superficies de respuesta en optimización, 164, 165, 167  
supuestos utilizados en calibración lineal, cálculos, 90, 110, 111

tabla de frecuencias, 22  
tabla de números aleatorios, 68, 74, 148, 201  
tendencia, prueba de significación para, 125  
teorema binominal, 124, 130  
teorema del limite central, 28, 122  
titulaciones complexométricas, ¿101?  
tolerancias, de materiales de vidrio y pesas, 7, 8  
transformaciones de gráficas de calibración curvas en lineales, 113, 116-118  
transmitancia, 34-35  
tratamientos, 148, 149, 153

uso secuencial de las pruebas de significación, 63

valor objetivo, en diagramas de control, 79-82, 84  
valor verdadero, 23, 25, 26, 28, 29, 40, 41  
valores anómalos, 2, 3, 49-51, 67, 78, 116, 123, 128, 129, 140  
valores críticos en pruebas estadísticas, 41, 42, 45-51, 55, 58, 70, 76, 77, 137, 143, 151, 152, 157, 160, 199-206  
valores de codificación previos a los cálculos, 21  
variable normal estándar, 143-145  
variación entre laboratorios, 71, 72, 76, 77  
varianza en la medición, 68-70  
varianza muestral, 68-71  
varianza, 22, 26, 33, 43, 47-49, 53, 54, 56, 63, 69, 70, 111, 128, 149, 151, 162, 198  
vector de pautas, 171, 172-173

Wold, S., 176-177

Yates, 58, 159  
Youden, W. J., 72, 73